

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: July 1, 2003, 05:18:35 ; Search time 585.641 Seconds

(without alignments)  
993.879 Million cell updates/sec

Title: US-10-053-662A-29

Perfect score: 20

Sequence: 1 tttaccagggagtgagaa 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_dln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hcg\_hum:\*  
40: em\_hcg\_mus:\*  
41: em\_hcg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	92.0	223	9 HSLAM110	U31187 Human lamn
2	18.4	92.0	2310	6 AK021648	AK021648 Homo sapi
3	18.4	92.0	3620	6 AX045567	AX045567 Sequence
4	18.4	92.0	3720	6 AX045565	AX045565 Sequence
5	18.4	92.0	4316	6 162750	162750 Sequence 14
6	18.4	92.0	4316	6 HSLAMB2TB	Z15009 H. sapiens m
7	18.4	92.0	5020	6 AX045563	AX045563 Sequence
8	18.4	92.0	5156	6 AX365737	AX365737 Sequence
9	18.4	92.0	5156	9 HSNICE	X73902 H. sapiens m
10	18.4	92.0	5200	6 AX045561	AX045561 Sequence
11	18.4	92.0	5200	6 162749	162749 Sequence 12
12	18.4	92.0	5200	6 HSLAMB2T	Z15008 H. sapiens m
13	18.4	92.0	98233	2 AC131536	AC131536 Rattus no
14	18.4	92.0	133936	9 HSDJ38304	AL109921 Human DNA
15	18.4	92.0	164084	9 AL354953	AL354953 Human DNA
16	18.4	92.0	170492	2 AL139013	AL139013 Homo sapi
17	17.4	87.0	31720	9 HSL196E3	Z68277 Human DNA s
18	17.4	87.0	80327	2 AC055889	AC055889 Homo sapi
19	17.4	87.0	145917	2 AC098602	AC098602 Rattus no
20	17.4	87.0	146576	8 AC074282	AC074282 Oryza sat
21	17.4	87.0	159382	8 AC116601	AC116601 Oryza sat
22	17.4	87.0	164682	9 AC006199	AC006199 Homo sapi
23	17.4	87.0	165080	2 AC117493	AC117493 Homo sapi
24	17.4	87.0	182843	2 AC095080	AC095080 Rattus no
25	17.4	87.0	196476	2 AC055890	AC055890 Homo sapi
26	17.4	87.0	199255	2 AC084195	AC084195 Homo sapi
27	17.4	85.0	175288	2 AC094555	AC094555 Rattus no
28	16.8	84.0	63892	2 AL450339	AL450339 Human DNA
29	16.8	84.0	64719	2 AC129496	AC129496 Homo sapi
30	16.8	84.0	67406	2 AC099897	AC099897 Mus muscu
31	16.8	84.0	74289	9 AC011813	AC011813 Homo sapi
32	16.8	84.0	84234	9 AC107975	AC107975 Homo sapi
33	16.8	84.0	94531	2 AC106626	AC106626 Rattus no
34	16.8	84.0	100000	9 AB020875	AB020875 Homo sapi
35	16.8	84.0	120978	9 AL732367	AL732367 Homo sapi
36	16.8	84.0	123071	2 AC121481	AC121481 Human DNA
37	16.8	84.0	123706	3 TBR2587	AL671259 Trypanoso
38	16.8	84.0	136788	2 AC117044	AC117044 Rattus no
39	16.8	84.0	144290	9 AC098866	AC098866 Homo sapi
40	16.8	84.0	146340	9 AP003479	AP003479 Homo sapi
41	16.8	84.0	148159	2 AC106365	AC106365 Rattus no
42	16.8	84.0	151453	2 AC118988	AC118988 Sus scrof
43	16.8	84.0	155655	2 AC118843	AC118843 Rattus no
44	16.8	84.0	153808	2 AC105845	AC105845 Rattus no
45	16.8	84.0	158561	2 AC114125	AC114125 Rattus no

#### ALIGNMENTS

RESULT 1  
HSLAM110  
LOCUS  
DEFINITION Human laminin gamma2 chain gene (LINC2), exon 10 and flanking  
sequences.  
ACCESSION U31187  
VERSION U31187.1 GI:1236307  
KEYWORDS  
SEGMENT  
SOURCE  
ORGANISM  
Homo sapiens.  
Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (sites)  
Kallunki,P., Sainio,K., Eddy,R., Byers,M., Kallunki,T., Sariola,H.,

**TITLE** Beck, K., Hirvonen, H., Shows, T. B. and Tryggvason, K.  
A truncated laminin chain homologous to the B2 chain: structure, spatial expression, and chromosomal assignment

**JOURNAL** J. Cell Biol. 119 (3), 679-693 (1992)

**PubMed** 13316279

**REFERENCE** 13316279

**AUTHORS** 2 (bases 1 to 223)  
Airenne, T., Haakana, H., Sainio, K., Kallunki, T., Kallunki, P., Sario, H., and Tryggvason, K.

**TITLE** Structure of the human laminin gamma 2 chain gene (LAMC2): alternative splicing with different tissue distribution of two transcripts

**JOURNAL** Genomics 32 (1), 54-64 (1996)

**MEDLINE** 96230326

**PUBMED** 87861121

**REFERENCE** 3 (bases 1 to 223)  
Airenne, T.

**AUTHORS** Direct Submission

**TITLE** Submitted (07-JUL-1995) Tomi Airenne, Biocenter Oulu and Department of Biochemistry, University of Oulu, Oulu, FIN-90570, Finland

**JOURNAL** Location/Qualifiers

**FEATURES** source  
1..223  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q25-32"  
<1..20  
/gene="LAMC2"  
21..203  
/number="10  
204..>223  
/gene="LAMC2"

**BASE COUNT** 45 a 65 c 62 g 51 t

**ORIGIN**

**Query Match** 92.0%; Score 18.4; DB 9; Length 223;  
Best Local Similarity 95.0%; Pred. No. 40;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**OY** 1 TGTACTCAGGCGATGAGAA 20  
||||| |||||||

**Db** 26 TGTATTACGGGATGAGAA 45

**RESULT 2**

**LOCUS** AK021648 2310 bp mRNA linear PRI 01-AUG-2002

**DEFINITION** Homo sapiens CDNA FLJ11586 fls, clone HEMBA1003720.

**ACCESSION** AK021648.1 GI:10432872

**VERSION** AK021648.1

**KEYWORDS** oligo capping; fls (full insert sequence);  
Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to mRNA, clone\_1lb:HEMBA1 clone:HEMBA1003720.

**SOURCE** Homo sapiens

**ORGANISM** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

**REFERENCE** 1  
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Magatsuma, M., Hosokiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, K., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K., and Iwayanagi, T.

**TITLE** NEDO human cDNA sequencing project

**JOURNAL** Unpublished

**REFERENCE** 2 (bases 1 to 2310)  
Isogai, T. and Otsuki, T.

**AUTHORS** Direct Submission

**TITLE** Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

**COMMENT** NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

**FEATURES** source  
1..2310  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEMBA1003720"  
/tissue\_type="whole embryo, mainly head"  
/clone\_1lb="HEMBA1"  
/dev\_stage="embryo, 10 weeks"  
/note="cloning vector: pME185FL3"

**BASE COUNT** 716 a 421 c 448 g 725 t

**ORIGIN**

**Query Match** 92.0%; Score 18.4; DB 9; Length 2310;  
Best Local Similarity 95.0%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**OY** 1 TGTACTCAGGCGATGAGAA 20  
||||| |||||||

**Db** 1786 TGTACTTACGGGATGAGAA 1805

**RESULT 3**

**LOCUS** AX045567 3620 bp DNA linear PAT 24-NOV-2000

**DEFINITION** Sequence 31 from Patent WO0066731.

**ACCESSION** AX045567

**VERSION** AX045567.1 GI:11344017

**KEYWORDS** human.

**SOURCE** Homo sapiens

**ORGANISM** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

**REFERENCE** 1 (bases 1 to 3620)  
Boutaud, A.

**AUTHORS** Recombinant laminin 5

**TITLE** Patent: WO 0066731-A 31 09-NOV-2000;

**JOURNAL** Biostatium, Inc. (US)

**FEATURES** source  
1..3620  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
<1..3519  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC17368.1"  
/db\_xref="GI:11344018"

**CDS** 1..3519  
/translat="TSRREYCDGCKGKSRQCFREHRLHGTGFGFCLNCDNTDGIHC  
ECKNGFYRHRERDRCLPCNCNKSGLSARCDNSGRCSKPGVTCARCDRCPCGHL  
TDAGCTQDRLRLDSKCDPAGIACPDGRCVCKPAYTGERCDGCRSGTYLDCGNP  
ECCTCCFCYGHASCRSSAEYSVHRITSTFHQDVGMVAOVQNSGPAKIQNSQRQD  
FSSAQLDPPVFPVAPAKFAGNOVYGGISLSDYVDVDRGRPSADVILEAGARIT  
APIMPLGTLIPGLRTYFRLNHEHSPNWSQSLSYFEYRLLRLMTLRITATGEY  
STGYIDNTYLISARVSGAPAVWOCIGKQFVKGOCODCAGSKRDSARLPGPGTC  
VPCNOGGGACDPDPCDYGSGDENPDICADPCIGFYNDPDPHRCGKPCNGPSGS  
VAPTEBEVYCNCPGVTGARGELCAGDTFCDPFEHGPYRRCQCCCNNTSPSASC  
NDRITLGRKLCIHTAGTGYCDQCKAGTFGDPLAVNPADKCRACNCPNGSPVPCRS  
DGTGCVKQFGPCNCEHGFSPACYNQVKTOMDPMOOLQMEALISKAOGDGVVP  
DELEGMOQAOLADILRDAOISEGASRSIGLOLAVRSQENSQSRDLDTKMTVE  
RYVALGSOVONRVDTHRLITOMOLSLAESRSIGLNTNIPADHYVPGKFSLOEA  
TRLASVYESASNMGLRRTEDYRKOALSLVRKALHGVSGSGSPGAVVQGIWEK  
LEKTSLAQDLREKTOAETEDRSYQHSRLDSVSPLOQYSDSFQYEKRRKQK  
ADLSLSLVTRHMDERKRTQKNGNKEEAOOLQNGKSGRESDSLVANLAKSAQ  
EALSNATIFYEYESILRNREFDQVNRKAEAEAMKRLSYISQVSDADKTQQA  
EALGSAADAOARANGAEALIEISSEIOEGISGICANRYADADLAMEGLAKLS  
EMREVEGLEERELEFDNMDAVOVVITAEQAVDTRAKNAGVTIDTNTLDTGLAHIM  
DQPLSVDEGLVLLEOKLSRAKTOINSQLRPMSELERARQGRHHLHLERSIDGIL

ADVKNLENIRDNLPFGCYNTQALEQO"  
BASE COUNT 959 a 862 c 1040 g 759 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 3620;  
Best Local Similarity 95.0%; Pred. No. 29;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGGGATGAGAA 20  
|||||  
Db 1228 TGTATTCAGGGGATGAGAA 1247

RESULT 4  
LOCUS AX045565 3720 bp DNA linear PAT 24-NOV-2000  
DEFINITION Sequence 29 from Patent WO0066731.  
ACCESSION AX045565  
VERSION AX045565.1 GI:11344015  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 3720)  
AUTHORS Boulaud, A.  
TITLE Recombinant laminin 5  
JOURNAL Patent: WO 0066731-A 29 09-NOV-2000;  
Bioslatum, Inc. (US)  
FEATURES  
source Location/Qualifiers  
1..3720  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
38..3619  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC17367.1"  
/db\_xref="GI:11344016"  
/translation="MPALMLGCGCLFSLILPAARATSRREVCDCNGKSRQCIFDRELH  
KOTNGEFCILNNDTGDIGCEKCNKNGYRHRERRCILPCNCSNGSSLSARDNSGRC  
SCKPGVTGARDRCILPGRHMLTDAGCTDORLDSKCDPCDPAIGPDAGRCVCKPA  
VTGERCDRCRGYINLDGNGPEGCTOCYGHASCRSAEYVAKITSTFRHVDGM  
KAVONGSPAKLQWQSRHODVSSAORLDPYFVAPAFELGQOVSYSQSLSDYRD  
RGRHPSADHVLIEGAGLRTAPLMLGKTLPCGLTKTFTPLNPNNSPOLSYF  
EYRLRLNLTALIRATYGEYSTGYIDVNTLISAPVSGAPAPWEGCICPYGKGF  
CODCASGYKRSARLPGPGTCLPCNCGGACDPDPTGDCYSGDENPDIECAACPIGFY  
NDPHDPRCKPCPCNHNPSGNCNCPGVGTGACELCAGYFEDPGEH  
GVPKPCQCCNNNDPSSAGNCDRLTGKLCINTAGIYCDCKAGYFEDPGEH  
ADKRCACNPNPSEPCVCKPGEPCGEGHAGFSPACYNQVOKTOMOFM  
OOLORMEALISKAGSGVPTDELEGRMOQAEOLLOPILDAQISEGASRLGQLA  
KYRSENYQSRHDLKMTVERVRAIGSOTQORVADTIRLITOMOLSAESASLGT  
NTPADHYVGPNGEKSIAOEATRLAESVESASNEQULREDEYSKALSLVRALH  
EGVSGSGSPGAVQGLVERLEKTSIAQULTRATAELEADSYKSLDLDSVS  
PLOGVSDSPFVEEAKRIKORADSLSLVTRHMDFTOKRLGNMKEADQILLDNGK  
SGREKSDQLSRANLAKSGRAOELSMGNATFEVSIILKNREPDLOYDNRAESEA  
MKRLSIOKYSADSDKTOQAEALGSAADQRAKNGAGEALFISSELEIGISILN  
EANYTADALAMEKSLASIKSEMYESELKELEFPTNMDAVQVYITEAKVDTRA  
KNAGYTDITLIDGLLHMDQPLSVDEEGVLLEOKLSRAKTIINSOLRPMSELE  
ERARQQRHLHLETSIDGILADVKNLENIRDNLPFGCYNTQALEQO"

sig-peptide  
BASE COUNT 969 a 900 c 1067 g 784 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 3720;  
Best Local Similarity 95.0%; Pred. No. 29;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGGGATGAGAA 20  
|||||  
Db 1328 TGTATTCAGGGGATGAGAA 1347

RESULT 5

162750  
LOCUS 162750 4316 bp DNA linear PAT 07-OCT-1997  
DEFINITION Sequence 14 from patent US 5660982.  
ACCESSION 162750  
VERSION 162750.1 GI:2480458  
KEYWORDS  
SOURCE  
ORGANISM  
Unclassified.  
1 (bases 1 to 4316)  
TRYGVASON, K., KALLUNKI, P. and PYKE, C.  
TITLE Laminin chains: diagnostic uses  
JOURNAL Patent: US 5660982-A 14 26-AUG-1997;  
Location/Qualifiers  
1..4316  
/organism="unknown"  
BASE COUNT 1158 a 1033 c 1226 g 899 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 4316;  
Best Local Similarity 95.0%; Pred. No. 28;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGGGATGAGAA 20  
|||||  
Db 1408 TGTATTCAGGGGATGAGAA 1427

RESULT 6  
LOCUS HSLAMB2TB 4316 bp mRNA linear PRI 27-MAR-1996  
DEFINITION H.sapiens mRNA for laminin.  
ACCESSION Z15009 S47253  
VERSION Z15009.1 GI:34231  
KEYWORDS laminin.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 4316)  
AUTHORS Trygvason, K.  
TITLE Direct Submission  
JOURNAL Submitted (27-AUG-1992) Trygvason K., Biocenter and University of  
Oulu, Biochemistry, Linnaana, Oulu, Finland, SF-90570  
2 (bases 1 to 4316)  
Kallunki, P., Sahlro, K., Eddy, R., Byers, M., Kallunki, T., Sariola, H.,  
Beck, K., Hirvonen, H., Shows, T. B. and Trygvason, K.  
A truncated laminin chain homologous to the B2 chain: structure,  
spatial expression, and chromosomal assignment  
J. Cell Biol. 119 (3), 679-693 (1992)  
J. Cell Biol. 93016279  
1383240

FEATURES  
source Location/Qualifiers  
1..4316  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/cell\_type="Fibroblasts"  
/cell\_type="Fibroblasts"  
118..3453  
/codon\_start=1  
/product="laminin"  
/protein\_id="CAA78729.1"  
/db\_xref="GI:34232"  
/db\_xref="SWISS-PROT:Q13753"  
/translation="MPALMLGCGCLFSLILPAARATSRREVCDCNGKSRQCIFDRELH  
KOTNGEFCILNNDTGDIGCEKCNKNGYRHRERRCILPCNCSNGSSLSARDNSGRC  
SCKPGVTGARDRCILPGRHMLTDAGCTDORLDSKCDPCDPAIGPDAGRCVCKPA  
VTGERCDRCRGYINLDGNGPEGCTOCYGHASCRSAEYVAKITSTFRHVDGM  
KAVONGSPAKLQWQSRHODVSSAORLDPYFVAPAFELGQOVSYSQSLSDYRD  
RGRHPSADHVLIEGAGLRTAPLMLGKTLPCGLTKTFTPLNPNNSPOLSYF  
EYRLRLNLTALIRATYGEYSTGYIDVNTLISAPVSGAPAPWEGCICPYGKGF  
CODCASGYKRSARLPGPGTCLPCNCGGACDPDPTGDCYSGDENPDIECAACPIGFY

CDS

NDPDRSCRCPCPCNCHNGFSCSVIETEEVVCNCPGVTGACELCADGYFEGDPGEH  
GVRRCQPCQCCQNDVPSASGNCDRITGRCCLICINHTAGIYCDQKAGYFGLPAPNP  
ADKRCACNCPNPGSEPCGSDGRCVCGPCGRCPCAGCPCACPCACPCACPCACPC  
CQLOREMLISAKOAGDGVPTPELEGRCQQAEOALQDILIRAOISEGASRSLGQILA  
KVRQSENYQSRHLDLAKMTVERVALGSOYOKRVYDTHLITOMULSLAESBASIGNT  
NIPADHYVGPNGFKSLAOEATRLAESVESASNEOQLTREDEDYSKOALSLVRLAH  
EYVGSQSPDPAVOVQGLVEKLEKTRSLAOQLTRATOETIADRSYOSHLLDNGS  
PLOGVSDSPVEAKRIKOKADSLSLVTRHMEFRCRKOKRKLKMKKEAQQLOLNGK  
SGRESDOLISANLAKSRAOALSMGNAPEVESEILKNPEPLQYDNRKAEAEBA  
MKRLSYISOKVSDASDKTQOARALGSAADQAKNGAGELIETSEEDQDGLTRLA  
EANTADGALAEKELASIKEMREVEBELKELEFDITMDAVOMVITEAKQVDTPLA  
KNAGVTIDTNTTDLGLHLMGM"

sig\_peptide  
repeat\_unit  
4021..4316  
118..183  
/rpt\_family="Human ALU"  
/rpt\_type=OTHER

polyA\_site  
BASE COUNT 1158 a 1033 c 1226 g 899 t  
ORIGIN

Query Match  
Best Local Similarity 92.0%; Score 18.4; DB 9; Length 4316;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGTACTCAGGGGATGAGAA 20  
||||| ||||||| |||||||  
1408 TGTATTCTAGGGGATGAGAA 1427

RESULT 7  
LOCUS AX045563 5020 bp DNA linear PAT 24-NOV-2000  
DEFINITION Sequence 27 from Patent W00066731.  
ACCESSION AX045563  
VERSION AX045563.1 GI:11344013  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 5020)  
AUTHORS Boulaud, A.  
TITLE Recombinant laminin 5  
JOURNAL Patent: WO 0066731-A 27 09-NOV-2000;  
Biotatium, Inc. (US)  
FEATURES  
source  
1..5020  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC17366.1"  
/db\_xref="GI:11344014"  
/translation="TSRRPYCDCKNGKSRQCIIRDELHROTNGFRCINCNNDTDTIC  
EKCKNGFTRHREDRKCLPCNCKSKSLKRCNDRSCRPVYFARCDRCPLGPHML  
TDAGCTQDQRLDKCDDPAGIAGPCAGRCVCKPAVTEGECRCRSYVLDGQNP  
ECCTQCFYGHASCSRSSAEVSHKITSFHDDVGMNAVONGSPAKIOMQRQDQV  
FSSAQRLDPVYVAPAKFIGNQVYSQSLSDYVDGRHRPASHLDVILEAGTGYE  
APIMPLGTLPCGLTKTYTERLNEHPSNNMSQQLSYFEYRGLRLNLTLLRATGRT  
STGYIDNTLISARVSGAPAPVBOCICPVYKQFOCDQCSGYKRSARLRPGCTC  
IPCNCQGGACDPTGDCYSGDENPDIIEADCPITFYDNDPHDRSCRCPCPCNNGSCS  
VLPTEVEVVCNCPGVTGARCELADGYFGDFEHGVRPCQPCQCCNPNPDSAGS  
NCDRLTGRCLKCIHNTAGIYCDQKAGYFGLPAPNPADKRCACNPNPGSPVCS  
DGTVCCKPGFGPCNCEHAFSCPCACYNQVCIOMDPMOOLQMEALISAKOAGDGVPT  
DELEGMOAEOALODILIRAOISEGASRSLGQLAVRSGENYOSRLDILAKTME  
PYRALGSOYONRDTLRLITOMULSLAESBASIGNTNIPADHYVGPNGFKSLAOE  
ATRLAESVESASNEOQLTREDEDYSKOALSLVRLAHESVGSQSPDPAVOVQGLVEK  
LEKTRSLAOQLTRATOETIADRSYOSHLLDNGSREKSDOLISANLAKSRAO  
ADSLSLVTRHMEFRCRKOKRKLKMKKEAQQLOLNGKSGRESDOLISANLAKSRAO  
EALISMGNAPEVESEILKNPEPLQYDNRKAEAEBA MKRLSYISOKVSDASDKTQO  
ARALGSAADQAKNGAGELIETSEEDQDGLTRLA EANTADGALAEKELASIKEMRE  
VEBELKELEFDITMDAVOMVITEAKQVDTPLA KNAGVTIDTNTTDLGLHLMGM

DOPLSVEDEGLVLLBOKLSRANKQINSOLRPMASELEBRARQORHLLLETSIDGIL  
ADVKNLEINRDLPPGCVNTQALQEO"

BASE COUNT 1329 a 1170 c 1333 g 1188 t  
ORIGIN

Query Match  
Best Local Similarity 92.0%; Score 18.4; DB 6; Length 5020;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGTACTCAGGGGATGAGAA 20  
||||| ||||||| |||||||  
1228 TGTATTCTAGGGGATGAGAA 1247

RESULT 8  
LOCUS AX365737 5156 bp DNA linear PAT 15-FEB-2002  
DEFINITION Sequence 130 from Patent W00200174.  
ACCESSION AX365737  
VERSION AX365737.1 GI:18697286  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1  
AUTHORS Wang, T., Wang, A., Skeiky, Y.A., Li, S.X., Kalos, M.D., Henderson, R.A.,  
McNeill, P.D., Fanger, N., Retter, M.W., Marnerakis, M., Fanger, G.R.,  
Wiedlick, T.S., Carter, D., Matanabe, Y., and Peckham, D.W.  
TITLE Compositions and methods for the therapy and diagnosis of lung  
JOURNAL cancer  
PATENT: WO 0200174-A 130 03-JAN-2002;  
CORIXA CORPORATION (US)  
FEATURES  
source  
1..5156  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 1351 a 1222 c 1377 g 1206 t  
ORIGIN

Query Match  
Best Local Similarity 92.0%; Score 18.4; DB 6; Length 5156;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGTACTCAGGGGATGAGAA 20  
||||| ||||||| |||||||  
1356 TGTATTCTAGGGGATGAGAA 1375

RESULT 9  
LOCUS HSNICE 5156 bp mRNA linear PRI 02-FEB-1994  
DEFINITION H.sapiens mRNA for niclein B2 chain.  
ACCESSION X73902  
VERSION X73902.1 GI:452754  
KEYWORDS Basement membrane; laminin; niclein; structural protein.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 5156)  
AUTHORS Vailly, J., Verrando, P., Champilaud, M.F., Gerecke, D., Wagman, D.W.,  
Baudoin, C., Aberdam, D., Burgess, R., Bauer, E. and Ortonne, J.P.  
TITLE The 100-kDa chain of niclein/kallinin is a laminin B2 chain variant  
JOURNAL Eur. J. Biochem. 219 (1-2), 209-218 (1994)  
MEDLINE 94139694  
PUBMED 8306988  
REFERENCE 2 (bases 1 to 5156)  
AUTHORS Vailly, J.  
TITLE Direct Submission  
JOURNAL submitted (01-JUL-1993) J. Vailly, Lab de Recherches  
Dermatologiques, Av De Valombrose Faculte de Medecine, 06107 Nice,  
Cedex 2, FRANCE



COMMENT Related sequence: Z15008.  
FEATURES Location/Qualifiers  
source 1..5156  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q25-31; In situ hybrid"  
/cell\_type="keratinocytes"  
66..3647  
/note="B2 chain; 100 kDa subunit"  
/codon\_start=1  
/product="nicotin"  
/protein\_id="CA52108.1"  
/db\_xref="GI:452755"  
/db\_xref="SWISS-PROT:Q13753"  
/translation="MPALMLGCCCLCSLLPAAATSRREVCDCNGKRCITFRELH  
RQNGNFRCLMNDNTDGIHECKNGFYRHRERDRLPCNCNSGSLASACDNGSGC  
SCKPGVIGARCDRLPGFHM/LTDAGCTODRLDSCDCPCAGIAGPDACVCKPA  
VTGERDCRSGYNIIDGNGNECTQCFCHSASCSASAYSVHKITSTFHQVDCM  
KAVORNGSPAKLQMSORHODVFSSAORLDPYFVAPAKFLGNOQVSGOISLFPYRD  
RGGHPSAHVYILEGAGLRITAPLMPILGKTLPCGLTKTTPRLMEHPSNMNSPOLSYF  
EYRLRLNLRALIRATYGERSTGYIDNVTLISRPVSGAFAPVVEQICVGTGKGF  
CODASGKRDASRLPGPTCTPCNCGGACDPTDGCYSGDENPDI ECDACPIGGE  
NDPDPSCRCPCPNHGFSCSVMPETEVEVCNCPGVTGARCELADGYDFDPEGH  
GVPAPCOPCCNNNVDSASGNCRLTGRCIKCIHNTAGIYCDCKAGYFEDPLAPN  
ADKRCACNKNMGSEPVGCRSDGTVCCKPGGNGCEHAGSCPACTYNQVITQDPM  
QOOLRMALISKAGGDDGVVDELEBMOQAEDLQDILDAQISGASASLQIQA  
KYRSQENSYSORLDDLMKTYERVALISOQYRVDRLHLLTQMLSLAESASLGNT  
NTPASDHYVGNFGRSLAQAETRLAESVESASMEQLTRETEDYSKQALSLVKALH  
EGVSGSGSPGAVOGLVEKLEKTSIAQOLTPRQAETREKQNLNKNKEAQQLQNGK  
RLQGVSDQSFVEEAKRIKQADLSLVTMHDEFRKQNLNKNKEAQQLQNGK  
SGREKSDQLSRANLAKSRAOELASMGNAFTYEVEISLKNREDFLOVQNKAAEEA  
MKRLSYISQKVSASDRTQOERALGSAADQAKNAGAGALISSEIPEQISLNL  
EANTADGALAMEKGLASLSKSEMEVEBELERKELEPTNDAYQVITTEAKQVTRA  
KNAGVTIDTNTLDGLHLMDOPLSYDEBELVLEQLISRAKQINSQALPMSSELE  
ERRAQOGRHLLLETSIDGLIADVKLENLINDNIPGCIYNTQALEQO"

BASE COUNT 1351 a 1222 c 1377 g 1206 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 5156;  
Best Local Similarity 95.0%; Pred. No. 28;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20  
|||||  
Db 1356 TGTATTTCAGGGGATGAGAA 1375

RESULT 10  
AX045561 5200 bp DNA linear PAT 24-NOV-2000  
LOCUS DEFINITION Sequence 25 from Patent W00066731.  
ACCESSION AX045561  
VERSION AX045561.1 GI:11344011  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 5200)  
AUTHORS Boutaud,A.  
TITLE Recombinant laminin 5  
JOURNAL Patent: WO 0066731-A 25 09-NOV-2000;  
BioStatum, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..5200  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
118..3699  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC17365.1"  
/db\_xref="GI:11344012"

/translation="MPALMLGCCCLCSLLPAAATSRREVCDCNGKRCITFRELH  
RQNGNFRCLMNDNTDGIHECKNGFYRHRERDRLPCNCNSGSLASACDNGSGC  
SCKPGVIGARCDRLPGFHM/LTDAGCTODRLDSCDCPCAGIAGPDACVCKPA  
VTGERDCRSGYNIIDGNGNECTQCFCHSASCSASAYSVHKITSTFHQVDCM  
KAVORNGSPAKLQMSORHODVFSSAORLDPYFVAPAKFLGNOQVSGOISLFPYRD  
RGGHPSAHVYILEGAGLRITAPLMPILGKTLPCGLTKTTPRLMEHPSNMNSPOLSYF  
EYRLRLNLRALIRATYGERSTGYIDNVTLISRPVSGAFAPVVEQICVGTGKGF  
CODASGKRDASRLPGPTCTPCNCGGACDPTDGCYSGDENPDI ECDACPIGGE  
NDPDPSCRCPCPNHGFSCSVMPETEVEVCNCPGVTGARCELADGYDFDPEGH  
GVPAPCOPCCNNNVDSASGNCRLTGRCIKCIHNTAGIYCDCKAGYFEDPLAPN  
ADKRCACNKNMGSEPVGCRSDGTVCCKPGGNGCEHAGSCPACTYNQVITQDPM  
QOOLRMALISKAGGDDGVVDELEBMOQAEDLQDILDAQISGASASLQIQA  
KYRSQENSYSORLDDLMKTYERVALISOQYRVDRLHLLTQMLSLAESASLGNT  
NTPASDHYVGNFGRSLAQAETRLAESVESASMEQLTRETEDYSKQALSLVKALH  
EGVSGSGSPGAVOGLVEKLEKTSIAQOLTPRQAETREKQNLNKNKEAQQLQNGK  
RLQGVSDQSFVEEAKRIKQADLSLVTMHDEFRKQNLNKNKEAQQLQNGK  
SGREKSDQLSRANLAKSRAOELASMGNAFTYEVEISLKNREDFLOVQNKAAEEA  
MKRLSYISQKVSASDRTQOERALGSAADQAKNAGAGALISSEIPEQISLNL  
EANTADGALAMEKGLASLSKSEMEVEBELERKELEPTNDAYQVITTEAKQVTRA  
KNAGVTIDTNTLDGLHLMDOPLSYDEBELVLEQLISRAKQINSQALPMSSELE  
ERRAQOGRHLLLETSIDGLIADVKLENLINDNIPGCIYNTQALEQO"

BASE COUNT 1364 a 1236 c 1392 g 1208 t  
ORIGIN

sig.peptide  
Query Match 92.0%; Score 18.4; DB 6; Length 5200;  
Best Local Similarity 95.0%; Pred. No. 28;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20  
|||||  
Db 1408 TGTATTTCAGGGGATGAGAA 1427

RESULT 11  
162749 5200 bp DNA linear PAT 07-OCT-1997  
LOCUS DEFINITION Sequence 12 from patent US 5660982.  
ACCESSION 162749  
VERSION 162749.1 GI:2480457  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5200)  
AUTHORS Trygvaason,K., kallunki,P. and Pyke,C.  
TITLE Laminin Chains: diagnostic uses  
JOURNAL Patent: US 5660982-A 12 26-AUG-1997;  
FEATURES Location/Qualifiers  
source 1..5200  
/organism="unknown"  
BASE COUNT 1364 a 1236 c 1392 g 1208 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 5200;  
Best Local Similarity 95.0%; Pred. No. 28;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20  
|||||  
Db 1408 TGTATTTCAGGGGATGAGAA 1427

RESULT 12  
HSLAMB2T 5200 bp mRNA linear PRI 27-MAR-1996  
LOCUS DEFINITION H.sapiens mRNA for laminin.  
ACCESSION Z15008 S47028  
VERSION Z15008.1 GI:34229  
KEYWORDS laminin.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



Consensus quality: 52180 bases at least Q30  
Consensus quality: 55537 bases at least Q20

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently consists of 48 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1046: contig of 1046 bp in length  
1047 1146: gap of unknown length  
1147 2583: contig of 1437 bp in length  
2584 2683: gap of unknown length  
2684 3721: contig of 1038 bp in length  
3722 3821: gap of unknown length  
3822 5429: contig of 1608 bp in length  
5430 5529: gap of unknown length  
5530 6756: contig of 1227 bp in length  
6757 6856: gap of unknown length  
6857 7887: contig of 1031 bp in length  
7888 9231: contig of 1244 bp in length  
9232 9331: gap of unknown length  
9332 10413: contig of 1082 bp in length  
10414 10513: gap of unknown length  
10514 12049: contig of 1536 bp in length  
12050 12149: gap of unknown length  
12150 13250: contig of 1101 bp in length  
13251 13350: gap of unknown length  
13351 14619: contig of 1269 bp in length  
14620 14719: gap of unknown length  
14720 15729: contig of 1010 bp in length  
15730 15829: gap of unknown length  
15830 16966: contig of 1137 bp in length  
16967 17066: gap of unknown length  
17067 18149: contig of 1083 bp in length  
18150 18249: gap of unknown length  
18250 19642: contig of 1393 bp in length  
19643 19742: gap of unknown length  
19743 21226: contig of 1484 bp in length  
21227 21326: gap of unknown length  
21327 22596: contig of 1270 bp in length  
22597 22696: gap of unknown length  
22697 23759: contig of 1063 bp in length  
23760 23859: gap of unknown length  
23860 25371: contig of 1512 bp in length  
25372 25471: gap of unknown length  
25472 26743: contig of 1272 bp in length  
26744 26843: gap of unknown length  
26844 28253: contig of 1410 bp in length  
28254 28353: gap of unknown length  
28354 30556: contig of 2203 bp in length  
30557 30656: gap of unknown length  
30657 32460: contig of 1804 bp in length  
32461 32560: gap of unknown length  
32561 33609: contig of 1049 bp in length  
33610 33709: gap of unknown length  
33710 35142: contig of 1433 bp in length  
35143 35242: gap of unknown length  
35243 36850: contig of 1608 bp in length  
36851 36950: gap of unknown length  
36951 38892: contig of 1942 bp in length  
38893 38992: gap of unknown length  
38993 40317: contig of 1325 bp in length  
40318 40417: gap of unknown length  
40419 41622: contig of 1205 bp in length  
41623 41722: gap of unknown length  
41723 43934: contig of 2212 bp in length  
43935 44034: gap of unknown length

44035 44035: contig of 1652 bp in length  
45687 45687: gap of unknown length  
45787 47335: contig of 1549 bp in length  
47336 47435: gap of unknown length  
47436 49695: contig of 2260 bp in length  
49696 49795: gap of unknown length  
49796 52881: contig of 3086 bp in length  
52882 52981: gap of unknown length  
52982 55039: contig of 2058 bp in length  
55040 55139: gap of unknown length  
55140 57295: contig of 2156 bp in length  
57296 57395: gap of unknown length  
57396 59052: contig of 1657 bp in length  
59053 59152: gap of unknown length  
59153 61795: contig of 2643 bp in length  
61796 61895: gap of unknown length  
61896 64192: contig of 2297 bp in length  
64193 64292: gap of unknown length  
64293 66929: contig of 2637 bp in length  
66930 67029: gap of unknown length  
67030 69843: contig of 2914 bp in length  
69844 70043: gap of unknown length  
70044 72183: contig of 2140 bp in length  
72184 72283: gap of unknown length  
72284 75162: contig of 2879 bp in length  
75163 75262: gap of unknown length  
75263 79581: contig of 4319 bp in length  
79582 79681: gap of unknown length  
79682 82554: contig of 2873 bp in length  
82555 82654: gap of unknown length  
82655 86943: contig of 4289 bp in length  
86944 87043: gap of unknown length  
87044 93641: contig of 6598 bp in length  
93642 93741: gap of unknown length  
93742 98233: contig of 4492 bp in length.

FEATURES  
source 1..98233  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-49E2"

BASE COUNT 24506 a 21274 c 21399 g 24747 t 6307 others  
ORIGIN

Query Match 92.0%; Score 18.4; DB 2; Length 98233;  
Best Local Similarity 95.0%; Pred. No. 20;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTTCACGAGGATGAGAA 20  
|||||||  
DB 30307 TGTTCACGAGGATGAGAA 30288

RESULT 14 133936 bp DNA linear PRI 30-SEP-2000  
HSDJ383J4  
LOCUS HSDJ383J4  
DEFINITION Human DNA sequence from clone Rps-383J4 on chromosome 1q24.1-24.3  
Contains part of a gene encoding a kelch motif containing protein,  
part of a novel gene encoding a protein similar to Aspartyl-tRNA  
synthetase, a putative novel gene, a 40S ribosomal protein S27  
(RPS27) pseudogene, 2 Cpg islands, ESTs, STSS and GSSs, complete  
sequence.

ACCESSION AL109921.21 GI:9588678  
KEYWORDS HTG: Aspartyl-tRNA synthetase; Cpg island; kelch-motif; RPS27.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 133936)  
AUTHORS Frankland,J.  
TITLE Direct Submision  
JOURNAL Submitted (25-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone



```

repeat_region /note="match: STS: Em:G05508"
16794..16949 /note="LMC/D repeat: matches 5497..5656 of consensus"
repeat_region 17170..17257 /note="LMC/D repeat: matches 5497..5656 of consensus"
repeat_region 17260..17359 /note="44 copies 2 mer at 73% conserved"
repeat_region 17475..17790 /note="50 copies 2 mer ta 75% conserved"
repeat_region 17826..18023 /note="LMC/D repeat: matches 5127..5403 of consensus"
repeat_region 17826..18023 /note="LMC/D repeat: matches 5127..5403 of consensus"
misc_feature /note="match: GSS: Em:A0621823"
18377..19732 /note="match: GSS: Em:A0621823"
repeat_region 19733..20035 /note="LMC/D repeat: matches 4372..5867 of consensus"
repeat_region 20036..20446 /note="Alusx repeat: matches 1..312 of consensus"
repeat_region 20449..20547 /note="LMC/D repeat: matches 3983..4372 of consensus"
repeat_region 20564..20972 /note="Alusg/x repeat: matches 193..291 of consensus"
repeat_region 21028..21531 /note="LMC/D repeat: matches 3522..3974 of consensus"
repeat_region 21851..22253 /note="LMC/D repeat: matches 2914..3451 of consensus"
repeat_region 22393..23023 /note="L1 repeat: matches 2110..2515 of consensus"
repeat_region 23235..23321 /note="L1M4 repeat: matches 3134..3809 of consensus"
repeat_region 23359..23459 /note="L1P43 repeat: matches 6056..6142 of consensus"
repeat_region 24526..24602 /note="L1M2 repeat: matches 6021..6122 of consensus"
repeat_region /note="MER33 repeat: matches 108..188 of consensus"

```

```

Query Match 92.0%; Score 18.4; DB 9; Length 133936;
Best Local Similarity 95.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 TGTACTCAGGAGATGAGAA 20
Db 43829 TGTACTCAGGAGATGAGAA 43848

RESULT 15
AL354953 164084 bp DNA linear PRI 16-NOV-2001
LOCUS Human DNA sequence from clone Rpl1-181K3 on chromosome 1, complete
DEFINITION sequence.
ACCESSION AL354953
VERSION AL354953.16 GI:16973821
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 164084)
REFERENCE
AUTHORS Tracey, A.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humangene@sanger.ac.uk; clone requests: clonerequests@sanger.ac.uk
On Nov 17, 2001 this sequence version replaced gi:16304444.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=

```

30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SwissProt; Tr, TrEMBL; Wp, WormPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep). This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>. Rpl1-181K3 is from the library RPCI-11.1 constructed by the group of Pliet de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

```

FEATURES
Source 1..164084
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="Rpl1-181K3"
/clone_11b="RPCI-11.1"
55965
misc_feature
/note="Random repeat. Forced join. Gap size estimated to
be approximately 80bp by restriction digest data."

```

```

BASE COUNT 46832 a 33365 c 35491 g 48396 t
ORIGIN

```

```

Query Match 92.0%; Score 18.4; DB 9; Length 164084;
Best Local Similarity 95.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 TGTACTCAGGAGATGAGAA 20
Db 133542 TGTACTCAGGAGATGAGAA 133561

```

```

Search completed: July 1, 2003, 07:22:57
Job time : 592.641 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 03:06:29 ; Search time 110.769 Seconds

(without alignments)  
406.611 Million cell updates/sec

Title: US-10-053-662A-29

Perfect score: 20

Sequence: 1 tftactcaggagatagaa 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: N\_Geneseq\_101002.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.4	92.0	580	22	AAH09229
2	18.4	92.0	2310	22	AAH15301
3	18.4	92.0	3582	24	AAO31145
4	18.4	92.0	3620	21	AAO31145
5	18.4	92.0	3720	21	AAO31145
6	18.4	92.0	4316	17	AAI13334
7	18.4	92.0	4316	24	AAI13334
8	18.4	92.0	5020	21	AAO31145
9	18.4	92.0	5152	20	AAZ24627

10	18.4	92.0	5156	21	AAO31145	Human lung cancer-
11	18.4	92.0	5156	24	AB149085	Human lung tumour
12	18.4	92.0	5200	17	AAI13323	Kallitri/Laminin 5
13	18.4	92.0	5200	21	AAO31145	Laminin gamma-2 ch
14	18.4	92.0	5200	24	AAI142910	Laminin gamma-2 ch
15	16.4	82.0	20066	24	AB152278	Human chemokine (C
16	16.4	82.0	14748	24	ABK51275	Human chemokine-2, C
17	16.4	82.0	17026	22	AAK70132	Human chemokine-2, C
18	16.4	82.0	22400	22	AAO31145	Human chemokine-2, C
19	15.8	79.0	793	20	AAO31145	Human chemokine-2, C
20	15.8	79.0	1358	20	AAO31145	Human chemokine-2, C
21	15.8	79.0	1745	22	AAK91274	Human chemokine-2, C
22	15.8	79.0	2392	24	AAO31145	Human chemokine-2, C
23	15.8	79.0	12261	22	AAO31145	Human chemokine-2, C
24	15.8	79.0	56632	22	AAO31145	Human chemokine-2, C
25	15.4	77.0	351	22	AAK57605	Human chemokine-2, C
26	15.4	77.0	1317	20	AAZ20164	Human chemokine-2, C
27	15.4	77.0	7591	22	AAK52887	Human chemokine-2, C
28	15.4	77.0	14540	22	AAK72853	Human chemokine-2, C
29	15.4	77.0	17596	22	AAK72853	Human chemokine-2, C
30	15.4	77.0	23307	23	AB19446	Human chemokine-2, C
31	15.4	77.0	31193	23	AB19446	Human chemokine-2, C
32	15.2	76.0	47	21	AAZ66815	Human chemokine-2, C
33	15.2	76.0	303	22	AAI34949	Human chemokine-2, C
34	15.2	76.0	425	22	AAI34949	Human chemokine-2, C
35	15.2	76.0	454	19	AAV60966	Human chemokine-2, C
36	15.2	76.0	499	22	AAK56847	Human chemokine-2, C
37	15.2	76.0	527	21	AAO31145	Human chemokine-2, C
38	15.2	76.0	531	22	AAI92349	Human chemokine-2, C
39	15.2	76.0	532	22	AAK61141	Human chemokine-2, C
40	15.2	76.0	565	23	AAO31145	Human chemokine-2, C
41	15.2	76.0	621	23	AAV5264	Human chemokine-2, C
42	15.2	76.0	696	19	AAV60967	Human chemokine-2, C
43	15.2	76.0	700	24	ABO5682	Human chemokine-2, C
44	15.2	76.0	814	23	AAO31145	Human chemokine-2, C
45	15.2	76.0	864	23	AAO31145	Human chemokine-2, C

## ALIGNMENTS

RESULT 1	
AAH09229/c	
ID	AAH09229 standard; cDNA: 580 BP.
XX	AAH09229;
XX	26-JUN-2001 (first entry)
XX	Human cDNA clone (3'-primer) SEQ ID NO:6064.
XX	Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX	Homo sapiens.
XX	EP1074617-A2.
XX	07-FEB-2001.
XX	28-JUL-2000; 2000EP-0116126.
XX	29-JUL-1999; 99JP-0248036.
XX	27-AUG-1999; 99JP-0300253.
XX	11-JAN-2000; 2000JP-0118776.
XX	02-MAY-2000; 2000JP-0183767.
XX	09-JUN-2000; 2000JP-0241899.
XX	(HELI-) HELIX RES INST.
XX	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	WPI; 2001-318749/34.



```
XX primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 3; SEQ ID 6064; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 580 BP; 122 A; 156 C; 126 G; 175 T; 1 other;
XX
Query Match          92.0%; Score 18.4; DB 22; Length 580;
Best Local Similarity 95.0%; Pred. No. 5.9;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TGTACTCAGGGGATGAGAA 20
    ||||| ||||| |||||
DB 524 TGTACTTATGGGGATGAGAA 505
XX
RESULT 2
AAH15301
ID AAH15301 standard; cDNA; 2310 BP.
XX
AC AAH15301;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:13453.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
```

```
DR WP1; 2001-318749/34.
XX
XX primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 13453; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2310 BP; 716 A; 421 C; 448 G; 725 T; 0 other;
XX
Query Match          92.0%; Score 18.4; DB 22; Length 2310;
Best Local Similarity 95.0%; Pred. No. 7;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TGTACTCAGGGGATGAGAA 20
    ||||| ||||| |||||
DB 1786 TGTACTTATGGGGATGAGAA 1805
XX
RESULT 3
AAD31145
ID AAD31145 standard; cDNA; 3582 BP.
XX
AC AAD31145;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human laminin gamma2 chain cDNA.
XX
KW Cancer; cell proliferation; integrin receptor; signalling pathway;
XX gene therapy; human; laminin gamma2; ss.
XX
OS Homo sapiens.
XX
PN WO200230465-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US32127.
XX
PR 12-OCT-2000; 2000US-239705P.
XX
PR 24-OCT-2000; 2000US-242812P.
XX
XX
PA (UNIV ) UNIV ROCHESTER.
XX
Key Location/Qualifiers
FT CDS 1..3582
FT /tag= a
FT /product= "laminin gamma2 chain"
```

```
XX Land H, Deleu L;
PI
XX WPI: 2002-416838/44.
DR P-PSDB; AAB4712.
XX Reducing proliferation of cancer cell, by inhibiting ligand binding to
PT integrin receptor on cell, reducing integrin-integrin interaction,
PT receptor clustering interaction or integrin-non-integrin protein
PI interaction
XX
XX Disclosure: Page 85-86; 148pp; English.
XX
CC The invention relates to a method of reducing proliferation of cancer
CC cells, comprising inhibiting ligand binding to integrin receptor on
CC cancer cells, where integrin receptor comprises integrin, reducing
CC integrin-integrin interaction, integrin receptor clustering interaction
CC or integrin-non-integrin protein interaction, reducing production of
CC integrin or ligand of integrin receptor by cancer cells, or
CC interfering with integrin-signalling pathway. The present sequence
CC is human laminin gamma2 chain cDNA.
XX
SQ Sequence 3582 BP; 938 A; 872 C; 1026 G; 746 T; 0 other;
XX
Query Match 92.0%; Score 18.4; DB 24; Length 3582;
Best Local Similarity 95.0%; Pred. No. 7.4;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TGTTCAGGCGGATGAGAA 20
DB 1291 TGTTCAGGCGGATGAGAA 1310
XX
RESULT 4
AAC83732
ID AAC83732 standard; cDNA; 3620 BP.
XX
AC AAC83732;
XX
DE 02-MAR-2001 (first entry)
XX
DE Human laminin 5 cDNA, SEQ ID NO: 31.
XX
XX Human; laminin 5; vulnereary; antiulcer; antiinflammatory; antidiabetic;
KW cell adhesion promoter; wound healing; ulcers; burn; skin graft;
KM periodontitis; gingivitis; Type I diabetes; angiogenesis regulation; ss.
XX
OS Homo sapiens.
XX
PA WO200066731-A2.
XX
PN 09-NOV-2000.
XX
PD 28-APR-2000; 2000WO-US11459.
XX
PE 30-APR-1999; 99US-0131720.
XX
PR 21-AUG-1999; 99US-0149738.
XX
PR 24-SEP-1999; 99US-0155945.
XX
PA (BIOS-) BIOSTATUM INC.
XX
XX
XX Boutaud A;
XX
XX WPI: 2000-687538/67.
XX
XX P-PSDB; AAB48471.
XX
XX Laminin 5-expressing cells, used to accelerate wound healing associated
PT with diabetic foot ulcers, venous ulcers, pressure sores, skin surgery,
PT burns, acute wounds and skin grafts -
XX
XX Claim 4; Page 204-209; 232pp; English.
XX
XX The present sequence encodes a laminin 5 chain polypeptide. Recombinant
```

```
CC laminin 5-expressing cells are used to accelerate wound healing,
CC especially diabetic foot ulcers, venous ulcers, pressure sores, skin
CC surgery, burns, acute wounds, skin grafts, corneal ulcerations,
CC gastro-intestinal ulcers, periodontitis, and gingivitis. They are also
CC used to improve the biocompatibility of medical devices, and to promote
CC cell adhesion to a surface. They can be used for the ex vivo treatment
CC of Type I diabetes. Laminin can also be used to regulate angiogenesis.
CC The cell line produces and secretes recombinant heterotrimeric laminin,
CC whereas prior art cell lines have been created that produce but do not
CC secrete only one or two chain laminins.
XX
SQ Sequence 3620 BP; 959 A; 862 C; 1040 G; 759 T; 0 other;
XX
Query Match 92.0%; Score 18.4; DB 21; Length 3620;
Best Local Similarity 95.0%; Pred. No. 7.4;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TGTTCAGGCGGATGAGAA 20
DB 1228 TGTTCAGGCGGATGAGAA 1247
XX
RESULT 5
AAC83731
ID AAC83731 standard; cDNA; 3720 BP.
XX
AC AAC83731;
XX
DE 02-MAR-2001 (first entry)
XX
DE Human laminin 5 cDNA, SEQ ID NO: 29.
XX
XX Human; laminin 5; vulnereary; antiulcer; antiinflammatory; antidiabetic;
KW cell adhesion promoter; wound healing; ulcers; burn; skin graft;
KM periodontitis; gingivitis; Type I diabetes; angiogenesis regulation; ss.
XX
OS Homo sapiens.
XX
PA WO200066731-A2.
XX
PN 09-NOV-2000.
XX
PD 28-APR-2000; 2000WO-US11459.
XX
PE 30-APR-1999; 99US-0131720.
XX
PR 21-AUG-1999; 99US-0149738.
XX
PR 24-SEP-1999; 99US-0155945.
XX
PA (BIOS-) BIOSTATUM INC.
XX
XX
XX Boutaud A;
XX
XX WPI: 2000-687538/67.
XX
XX P-PSDB; AAB48470.
XX
XX Laminin 5-expressing cells, used to accelerate wound healing associated
PT with diabetic foot ulcers, venous ulcers, pressure sores, skin surgery,
PT burns, acute wounds and skin grafts -
XX
XX Claim 4; Page 195-200; 232pp; English.
XX
XX The present sequence encodes a laminin 5 chain polypeptide. Recombinant
CC laminin 5-expressing cells are used to accelerate wound healing,
CC especially diabetic foot ulcers, venous ulcers, pressure sores, skin
CC surgery, burns, acute wounds, skin grafts, corneal ulcerations,
CC gastro-intestinal ulcers, periodontitis, and gingivitis. They are also
CC used to improve the biocompatibility of medical devices, and to promote
CC cell adhesion to a surface. They can be used for the ex vivo treatment
CC of Type I diabetes. Laminin can also be used to regulate angiogenesis.
CC The cell line produces and secretes recombinant heterotrimeric laminin,
CC whereas prior art cell lines have been created that produce but do not
CC secrete only one or two chain laminins.
XX
```

SO Sequence 3720 BP; 969 A; 900 C; 1067 G; 784 T; 0 other;

Query Match 92.0%; Score 18.4; DB 21; Length 3720;

Best Local Similarity 95.0%; Pred. No. 7.5;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTTCAGGAGATGAGAA 20

DB 1328 TGTTCAGGAGATGAGAA 1347

RESULT 6

AA13324

ID AA13324 standard; cDNA; 4316 BP.

AC AA13324;

DT 13-NOV-1996 (first entry)

DE Kallinin/laminin 5 gamma-2 chain alternative coding sequence.

XX Kallinin; laminin; epidermolysis bullosa; junctional; probe;

KM detection; inhibit; monitor; malignancy; ss.

XX Homo sapiens.

OS Homo sapiens.

FN Key Location/Qualifiers

FT CDS 118..3453

FT /tag= b

FT /note= "Kallinin/laminin 5 gamma-2 chain"

PN MO9610646-A1.

XX 11-APR-1996.

PF - 04-OCT-1995; 95MO-EP03918.

PR 04-OCT-1994; 94US-0317450.

PA (TRYG/) TRYGVASON K.

PI Kallunki P, Pyke C, Trygvason K;

DR WPI; 1996-209366/21.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

XX P-PSDB; AAR91428.

ID AAL42911 standard; cDNA; 4316 BP.

XX AAL42911;

XX 05-AUG-2002 (first entry)

DE Laminin gamma-2 chain cDNA sequence 2.

XX Laminin gamma-2; gene; ss; cancer; laminin gamma-2 chain inhibition;

KM carcinoma; inhibition; anti-gamma-2 chain antibody;

XX epithelial cell adhesion; laminin-5.

XX Undifferentiated.

OS Undifferentiated.

FN Key Location/Qualifiers

FT CDS 118..3453

FT /tag= a

FT /product= "Laminin gamma-2 chain 2"

FT sig\_peptide 118..183

FT mat\_peptide 184..3450

FT /tag= c

FT /note= "Mature laminin gamma-2 chain 2"

PN US2002052307-A1.

XX 02-MAY-2002.

PF 08-JAN-2001; 2001US-0756071.

PR 07-JAN-2000; 2000US-175005P.

PR 04-OCT-1994; 94US-0317450.

PR 18-FEB-1997; 97US-0800593.

PR 15-SEP-2000; 2000US-0663147.

PA (TRYG/) TRYGVASON K.

PI (KALL/) KALLUNKI P.

XX (PYKE/) PYKE C.

XX Trygvason K, Kallunki P, Pyke C;

DR WPI; 2002-434824/46.

XX P-PSDB; AAO14993.

XX P-PSDB; AAO14993.

XX P-PSDB; AAO14993.

XX P-PSDB; AAO14993.

XX P-PSDB; AAO14993.

XX P-PSDB; AAO14993.

XX P-PSDB; AAO14993.

XX P-PSDB; AAO14993.

XX P-PSDB; AAO14993.

XX P-PSDB; AAO14993.

XX P-PSDB; AAO14993.

SO Sequence 4316 BP; 1158 A; 1033 C; 1226 G; 899 T; 0 other;

Query Match 92.0%; Score 18.4; DB 24; Length 4316;

Best Local Similarity 95.0%; Pred. No. 7.6;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTTCAGGAGATGAGAA 20

DB 1408 TGTTCAGGAGATGAGAA 1427

RESULT 8

AA142911



XX Wang T, Fan L;  
PI WPI; 2000-628399/60.  
XX  
XX  
XX Isolated polypeptide comprising an immunogenic portion of a lung tumor  
PT protein is used for detecting and monitoring progression of lung cancer  
PT in a patient -  
XX  
XX  
XX Claim 25a; Page 152-153; 261pp; English.  
XX  
XX This invention describes a novel isolated polypeptide (I) which  
CC comprising an immunogenic portion of a lung tumor protein or variant (P2)  
CC which have cytostatic activity. The polypeptides and polynucleotides are  
CC used in compositions and vaccines to inhibit the development of cancer,  
CC especially lung cancer, in a patient. Methods described in the invention  
CC can be used to monitor the progression of a cancer by carrying out the  
CC detection at subsequent time points and comparing the results from the  
CC different time points. CD4+ and/or CD8+ T-Cells isolated from a patient  
CC are treated with P2, polynucleotides encoding P2 or antigen presenting  
CC cells expressing P2 and then administered to the patient to inhibit  
CC development of cancer.  
XX  
XX Sequence 5156 BP; 1351 A; 1222 C; 1377 G; 1206 T; 0 other;  
SQ  
Query Match 92.0%; Score 18.4; DB 21; Length 5156;  
Best Local Similarity 95.0%; Pred. No. 7.8;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TGTACTCAGGAGATGAGAA 20  
DB 1356 TGTATTCAGGAGATGAGAA 1375  
RESULT 11  
ABL49085  
ID ABL49085 standard; cDNA; 5156 BP.  
XX  
XX ABL49085;  
AC  
XX  
XX 01-MAY-2002 (first entry)  
DT  
XX  
XX Human lung tumour cDNA sequence for contig 20 SEQ ID NO:130.  
DE  
XX  
XX Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;  
KW immune response; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200200174-A2.  
PN  
XX  
XX 03-JAN-2002.  
PD  
XX  
XX 28-JUN-2001; 2001WO-US21065.  
PF  
XX  
XX 28-JUN-2000; 2000US-0606421.  
PR  
XX  
XX 02-AUG-2000; 2000US-0630940.  
PR  
XX  
XX 21-AUG-2000; 2000US-0643597.  
PR  
XX  
XX 15-SEP-2000; 2000US-0662786.  
PR  
XX  
XX 09-OCT-2000; 2000US-0685696.  
PR  
XX  
XX 12-DEC-2000; 2000US-0735705.  
PR  
XX  
XX 07-MAY-2001; 2001US-0850716.  
XX  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX  
XX Wang T, Wang A, Skelky YAM, Li SX, Kalos MD, Henderson RA;  
PI McNeill PD, Fanger N, Retter MW, Marnierakis M, Fanger GR;  
PI Vedick TS, Carter D, Watanabe Y, Peckham DW;  
XX  
XX WPI; 2002-090513/12.  
DR  
XX  
XX Polynucleotides encoding lung tumor polypeptides, useful for treating  
PT lung cancer or stimulating an immune response -

XX Example 3; Page 236-237; 374pp; English.  
PS  
XX  
XX The present invention describes human lung tumour proteins. Human lung  
CC tumour proteins and polynucleotides have cytostatic and immunostimulant  
CC activities, and can be used in vaccine production. Compositions  
CC comprising the lung tumour proteins, polynucleotides, antibodies,  
CC fusion proteins, T cell populations, or antigen presenting cells that  
CC express the lung tumour proteins are useful for treating lung cancer or  
CC stimulating an immune response. ABL49085 to ABL49300 and ABL74946 to  
CC ABL75070 represent sequences used in the exemplification of the present  
CC invention.  
XX  
XX Sequence 5156 BP; 1351 A; 1222 C; 1377 G; 1206 T; 0 other;  
SQ  
Query Match 92.0%; Score 18.4; DB 24; Length 5156;  
Best Local Similarity 95.0%; Pred. No. 7.8;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TGTACTCAGGAGATGAGAA 20  
DB 1356 TGTATTCAGGAGATGAGAA 1375  
RESULT 12  
AAR13323  
ID AAR13323 standard; cDNA; 5200 BP.  
XX  
XX AAR13323;  
AC  
XX  
XX 13-NOV-1996 (first entry)  
DT  
XX  
XX Kalinin/Laminin 5 gamma-2 chain coding sequence.  
DE  
XX  
XX Kalinin; laminin; epidermolysis bullosa; junctional; probe;  
KW detection; inhibit; monitor; malignancy; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH CDS 118..3699  
FT /tag=a  
FT /product= kalinin/Laminin-5-gamma-2\_chain  
XX  
XX WO9610646-A1.  
PN  
XX  
XX 11-APR-1996.  
PD  
XX  
XX 04-OCT-1995; 95WO-EP03918.  
PF  
XX  
XX 04-OCT-1994; 94US-0317450.  
PR  
XX  
XX (TRYG/) TRYGGVASON K.  
PA  
XX  
XX kallunk1 P, Pyke C, Tryggvason K;  
PI  
XX  
XX WPI; 1996-209366/21.  
DR  
XX  
XX P-PSDB; AAR91427.  
PR  
XX  
XX Detection of kalinin or laminin 5 expression in cells - useful to  
PT detect, monitor and inhibit the invasive growth of cell in tissue,  
PT partic. malignant tissue  
XX  
XX Disclosure; Fig 4A; 37pp; English.  
PS  
XX  
XX The present sequence is the full cDNA (Genbank Z15008) encoding the  
CC kalinin/laminin 5 gamma-2 chain (R91427). The gamma-2 chain is of  
CC importance to patients suffering from epidermolysis bullosa, esp. the  
CC junctional form (JEB). Probes and antisense gamma-2 sequences derived  
CC from this sequence can be used to detect, monitor and inhibit the  
CC invasive growth of cells in tissue, partic. malignant tissue.  
XX  
XX Sequence 5200 BP; 1364 A; 1236 C; 1392 G; 1208 T; 0 other;  
SQ

Query Match 92.0%; Score 18.4; DB 17; Length 5200;  
Best Local Similarity 95.0%; Pred. No. 7.8;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGCGATGAGAA 20  
DB 1408 TGTATTCTCAGGCGATGAGAA 1427

RESULT 13  
AAC83729  
ID AAC83729 standard; CDNA: 5200 BP.

AC AAC83729;  
DT 02-MAR-2001 (first entry)  
DE Human laminin 5 CDNA, SEQ ID NO: 25.

Human laminin 5; vulnereary; antiulcer; antiinflammatory; antidiabetic;  
cell adhesion promoter; wound healing; ulcers; burn; skin graft;  
periodontitis; gingivitis; Type I diabetes; angiogenesis regulation; ss.  
XX Homo sapiens.

PN WO200066731-A2.

PD 09-NOV-2000.

PF 28-APR-2000; 2000WO-US11459.

PR 30-APR-1999; 99US-0131720.

PR 21-AUG-1999; 99US-0149738.

PR 24-SEP-1999; 99US-0155945.

PA (BIOS-) BIOSTATUM INC.

PI Boutaud A;

DR WPI: 2000-687538/67.

DR P-PSDB: AAB48468.

PT Laminin 5-expressing cells, used to accelerate wound healing associated  
PT with diabetic foot ulcers, venous ulcers, pressure sores, skin surgery,  
PT burns, acute wounds and skin grafts -

PS Claim 4; Page 175-181; 232pp; English.

CC The present sequence encodes a laminin 5 chain polypeptide. Recombinant  
CC laminin 5-expressing cells are used to accelerate wound healing,  
CC especially diabetic foot ulcers, venous ulcers, pressure sores, skin  
CC surgery, burns, acute wounds, skin grafts, corneal ulcerations,  
CC gastro-intestinal ulcers, periodontitis, and gingivitis. They are also  
CC used to improve the biocompatibility of medical devices, and to promote  
CC cell adhesion to a surface. They can be used for the ex vivo treatment  
CC of Type I diabetes. Laminin can also be used to regulate angiogenesis.  
CC The cell line produces and secretes recombinant heterotrimeric laminin,  
CC whereas prior art cell lines have been created that produce but do not  
CC secrete only one or two chain laminins.

SQ Sequence 5200 BP; 1364 A; 1236 C; 1392 G; 1208 T; 0 other;

Query Match 92.0%; Score 18.4; DB 21; Length 5200;

Best Local Similarity 95.0%; Pred. No. 7.8;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGCGATGAGAA 20  
DB 1408 TGTATTCTCAGGCGATGAGAA 1427

RESULT 14

AAL42910  
ID AAL42910 standard; CDNA: 5200 BP.

AC AAL42910;

DT 05-AUG-2002 (first entry)

DE Laminin gamma-2 chain CDNA sequence 1.

XX Laminin gamma-2; gene; ss; cancer; laminin gamma-2 chain inhibition;  
XX carcinogen inhibition; anti-gamma-2 chain antibody;  
XX epithelial cell adhesion; laminin-5.

OS Unidentified.

FI Key Location/Qualifiers  
FI CDS 118..3699  
FI /tag= a  
FI /product= "Laminin gamma-2 chain 1"

FI sig\_peptide 118..183  
FI /tag= b  
FI mat\_peptide 184..3696  
FI /tag= c  
FI /note= "Mature laminin gamma-2 chain 1"

US2002052307-A1.

PD 02-MAY-2002.

PF 08-JAN-2001; 2001US-0756071.

PR 07-JAN-2000; 2000US-175005P.

PR 04-OCT-1994; 94US-0317450.

PR 18-FEB-1997; 97US-0800593.

PR 15-SEP-2000; 2000US-0663147.

PA (TRYG/) TRYGIVASON K.

PA (KALL/) KALLUNKI P.

PA (PYKE/) PYKE C.

PI Tryggevason K, Kallunki P, Pyke C;

DR WPI: 2002-434824/46.  
DR P-PSDB: AAO14992.

PT Modulating laminin 5 gamma 2 chain interactions of invasive carcinogens  
PT for treating cancers and promoting attachment of cultured cells in  
PT vitro -

PS Example 1; Fig 4; 51pp; English.

CC The invention comprises a method of inhibiting the laminin gamma-2 chain  
CC interactions of invasive carcinogens with surrounding tissues - by using  
CC anti-gamma-2 chain antibodies to inhibit the gamma-2 chain biological  
CC activity of the invasive carcinogens. The invention also comprises a  
CC method for promoting adhesion of epithelial cells by exposing the cells  
CC to intact laminin-5 molecules. The first method of the invention is  
CC useful for preventing gamma 2 chain interactions of invasive carcinogens  
CC with surrounding tissues. The second method of the invention is useful  
CC for promoting adhesion of cultured epithelial cells. The present CDNA  
CC sequence encodes a laminin gamma-2 chain.

SQ Sequence 5200 BP; 1364 A; 1236 C; 1392 G; 1208 T; 0 other;

Query Match 92.0%; Score 18.4; DB 24; Length 5200;

Best Local Similarity 95.0%; Pred. No. 7.8;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGCGATGAGAA 20  
DB 1408 TGTATTCTCAGGCGATGAGAA 1427

RESULT 15  
ABL52278  
ID ABL52278 standard; DNA; 20066 BP.  
XX  
AC ABL52278;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human chemokine (C-C motif) receptor 6 (CCR6) gene SEQ ID NO.1.  
XX  
KM Human: chemokine (C-C motif) receptor 6; chemokine receptor 6; CCR6;  
KM receptor; polymorphic; single nucleotide polymorphism; SNP; genotyping;  
KM haplotyping; antipsoriatic; gene therapy; antisense gene therapy;  
KM psoriasis; gene; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT misc\_feature 1060  
FT /tag= a  
FT /note= "PS1: polymorphic base A or G"  
FT 1134  
FT /tag= b  
FT /note= "PS2: polymorphic base G or A"  
FT 1138  
FT /tag= c  
FT /note= "PS3: polymorphic base C or T"  
FT 1350  
FT /tag= d  
FT /note= "PS4: polymorphic base G or A"  
FT 1352  
FT /tag= e  
FT /note= "PS5: polymorphic base A or G"  
FT 1687  
FT /tag= f  
FT /note= "PS6: polymorphic base G or A"  
FT 15266  
FT /tag= g  
FT /note= "PS7: polymorphic base C or T"  
FT 16119  
FT /tag= h  
FT /note= "PS8: polymorphic base T or C"  
XX  
PN WO200226764-A2.  
XX  
PD 04-APR-2002.  
XX  
PF 24-SEP-2001; 2001MO-US29823.  
XX  
PR 26-SEP-2000; 2000US-235705P.  
XX  
PA (GENA-) GENAISSANCE PHARM INC.  
XX  
PI Chew A, Choi JY, Koshy B.  
XX  
DR WPI; 2002-394235/42.  
XX  
PT New genetic variants of chemokine (C-C) motif receptor 6, useful for  
PT therapeutic purposes and for expressing CCR6 protein useful in  
PT identifying drugs to treat psoriasis  
XX  
PS Claim 14; Fig 1; 73pp; English.  
XX  
CC The present invention describes an isolated polynucleotide (I) comprising  
CC a nucleotide sequence which is a polymorphic variant of a reference  
CC sequence for chemokine (C-C motif) receptor 6 (CCR6) gene (see ABL52278)  
CC or its fragment, or a polymorphic variant of a reference sequence for a  
CC CCR6 cDNA (see ABL52279) or its fragment. Also described is are methods  
CC for haplotyping and genotyping the CCR6 gene of an individual, and a  
CC method for identifying (II) an association between a trait and a  
CC haplotype or haplotype pair of CCR6 gene, by comparing the frequency of  
CC the haplotypes 1-13 or haplotype pair in a population exhibiting the  
CC trait with the frequency of the haplotype or haplotype pair in a

CC reference population. (II) has antipsoriatic activity, and can be used in  
CC gene therapy and antisense gene therapy. (II) is useful for identifying  
CC an association between a trait such as a clinical response to a drug  
CC targeting CCR6 and a haplotype or haplotype pair of CCR6 gene. Methods  
CC from the present invention have applicability in developing diagnostic  
CC tests and therapeutic treatments for psoriasis. (I) is useful for  
CC studying the expression and function of CCR6 and expressing CCR6 proteins  
CC for use in screening for candidate drugs to treat diseases related to  
CC CCR6 activity. Establishing the CCR6 haplotype or haplotype pair of an  
CC individual is useful for improving the efficiency and reliability of  
CC several steps in the discovery and development of drugs for treating  
CC diseases associated with CCR6 activity, such as psoriasis. The present  
CC sequence represents the CCR6 gene, where the polymorphic bases in the  
CC sequence are indicated using ambiguity bases.  
XX

SQ Sequence 20066 BP; 5198 A; 4490 C; 4450 G; 5920 T; 8 other;

Query Match 82.0%; Score 16.4; DB 24; Length 20066;

Best Local Similarity 94.4%; Pred. No. 1e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GTTACTCAGGCGATGAGA 19  
1 |||||

DB 8430 GATACCTCAGGCGATGAGA 8447

Search completed: July 1, 2003, 06:34:03  
Job time : 112.769 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 06:21:32 : Search time 26.6667 seconds  
(without alignments)  
230.008 Million cell updates/sec

Title: US-10-053-662a-29

Perfect score: 20

Sequence: 1 tggtactcagggatgagaa 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: Issued\_Patents\_NA:\*  
2: /cgn2\_6/ptodata/1/1na/5A\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/1na/5B\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/1na/6A\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/1na/6B\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/1na/PCrUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/1na/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	92.0	4316	1 US-08-317-450B-14	Sequence 14, Appl
2	18.4	92.0	4316	3 US-08-800-593-14	Sequence 14, Appl
3	18.4	92.0	5156	4 US-09-643-597-130	Sequence 130, App
4	18.4	92.0	5200	1 US-08-317-450B-12	Sequence 12, Appl
5	18.4	92.0	5200	3 US-08-800-593-12	Sequence 12, Appl
6	15.2	76.0	454	4 US-09-381-150A-1	Sequence 1, Appl
7	15.2	76.0	696	4 US-09-381-150A-2	Sequence 2, Appl
8	15.2	76.0	1810	1 US-07-913-107-1	Sequence 1, Appl
9	15.2	76.0	1810	1 US-08-459-201-1	Sequence 1, Appl
10	15.2	76.0	1810	1 US-08-281-248-1	Sequence 1, Appl
11	15.2	76.0	19227	3 US-09-090-793-13	Sequence 13, Appl
12	15.2	76.0	40138	3 US-09-080-793-12	Sequence 12, Appl
13	15.2	76.0	176373	3 US-09-128-153-17	Sequence 17, Appl
14	14.8	74.0	300	1 US-08-765-081-12	Sequence 12, Appl
15	14.8	74.0	300	3 US-09-098-082-12	Sequence 12, Appl
16	14.8	74.0	300	5 PCT-US95-06994-14	Sequence 14, Appl
17	14.8	74.0	485	4 US-09-484-970B-120	Sequence 120, App
18	14.8	74.0	2589	1 US-08-653-740-6	Sequence 6, Appl
19	14.8	74.0	2589	2 US-09-073-594-6	Sequence 6, Appl
20	14.8	74.0	2589	3 US-09-275-925-6	Sequence 6, Appl
21	14.8	74.0	17949	4 US-09-087-465-3	Sequence 3, Appl
22	14.8	74.0	43360	4 US-09-453-702B-206	Sequence 206, App
23	14.8	74.0	45325	4 US-09-453-702B-261	Sequence 261, App
24	14.8	74.0	169998	4 US-09-676-610B-24	Sequence 24, App
25	14.4	72.0	673	4 US-09-328-111-303	Sequence 303, App
26	14.4	72.0	1322	2 US-08-805-965-2	Sequence 2, Appl
27	14.4	72.0	1502	2 US-08-805-965-6	Sequence 6, Appl

#### ALIGNMENTS

28	14.4	72.0	2175	4 US-08-914-999-3	Sequence 3, Appl
29	14.2	71.0	263	3 US-09-004-113-8	Sequence 8, Appl
30	14.2	71.0	541	4 US-09-404-879A-11	Sequence 11, Appl
31	14.2	71.0	773	4 US-09-227-357-47	Sequence 47, Appl
32	14.2	71.0	1001	4 US-09-641-638-465	Sequence 465, App
33	14.2	71.0	1001	4 US-09-641-638-466	Sequence 466, App
34	14.2	71.0	1002	4 US-09-641-638-595	Sequence 595, App
35	14.2	71.0	1491	1 US-08-137-614A-3	Sequence 3, Appl
36	14.2	71.0	1491	2 US-08-768-301-1	Sequence 1, Appl
37	14.2	71.0	1701	4 US-09-411-628-5	Sequence 5, Appl
38	14.2	71.0	1861	4 US-09-129-668-7	Sequence 7, Appl
39	14.2	71.0	1897	1 US-08-184-632-1	Sequence 1, Appl
40	14.2	71.0	1981	4 US-08-981-392-26	Sequence 26, Appl
41	14.2	71.0	2065	4 US-09-129-668-5	Sequence 5, Appl
42	14.2	71.0	2178	3 US-09-034-916-1	Sequence 1, Appl
43	14.2	71.0	2265	3 US-08-906-865-2	Sequence 2, Appl
44	14.2	71.0	2265	4 US-09-129-668-2	Sequence 2, Appl
45	14.2	71.0	2457	4 US-08-872-757-1	Sequence 1, Appl

RESULT 1  
US-08-317-450B-14  
Sequence 14, Application US/08317450B  
Patent No. 5660982  
GENERAL INFORMATION:  
APPLICANT: Tryggvason, Karl  
APPLICANT: Kallunki, Pekka  
APPLICANT: Pyke, Charles  
TITLE OF INVENTION: Laminin Chains: Diagnostic and  
TITLE OF INVENTION: Therapeutic Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: BANNER & ALLEGRETTI, LTD.  
STREET: Ten South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,450B  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Chao, Mark  
REGISTRATION NUMBER: 37,293  
REFERENCE/DOCKET NUMBER: 94,778  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4316 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 118..183  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 118..3453  
FEATURE:  
NAME/KEY: repeat\_unit  
LOCATION: 4021..4316

OTHER INFORMATION: /rpt\_type="other"  
OTHER INFORMATION: /rpt\_family="HUMAN ALU"  
FEATURE:  
NAME/KEY: polyA\_site  
LOCATION: 4296  
US-08-317-450B-14

Query Match 92.0%; Score 18.4; DB 1; Length 4316;  
Best Local Similarity 95.0%; Pred. No. 0.64;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGCGATGAGAA 20  
||||| |||||||  
DB 1408 TGTATTTCAGGCGATGAGAA 1427

RESULT 2  
US-08-800-593-14  
Sequence 14, Application US/08800593  
Patent No. 6143505  
GENERAL INFORMATION:  
APPLICANT: Trygvason, Karl  
APPLICANT: Kallunki, Pekka  
APPLICANT: Pyke, Charles  
TITLE OF INVENTION: Laminin Chains: Diagnostic and  
TITLE OF INVENTION: Therapeutic Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
STREET: 300 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,593  
FILING DATE: 18-FEB-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/317,450  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Chao, Mark  
REGISTRATION NUMBER: 37,293  
REFERENCE/DOCKET NUMBER: 94,778-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-913-0001  
TELEFAX: 312-913-0002  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4316 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 118..183  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 118..3453  
FEATURE:  
NAME/KEY: repeat\_unit  
LOCATION: 4021..4316  
OTHER INFORMATION: /rpt\_type="other"  
OTHER INFORMATION: /rpt\_family="HUMAN ALU"  
FEATURE:

NAME/KEY: polyA\_site  
LOCATION: 4296  
US-08-800-593-14

Query Match 92.0%; Score 18.4; DB 3; Length 4316;  
Best Local Similarity 95.0%; Pred. No. 0.64;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGCGATGAGAA 20  
||||| |||||||  
DB 1408 TGTATTTCAGGCGATGAGAA 1427

RESULT 3  
US-09-643-597-130  
Sequence 130, Application US/09643597  
Patent No. 6426072  
GENERAL INFORMATION:  
APPLICANT: Wang, Tongtong  
APPLICANT: Fan, Liqun  
APPLICANT: Kalos, Michael D.  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: Hosken, Nancy  
APPLICANT: Fanger, Gary R.  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aljun  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Henderson, Robert A.  
APPLICANT: McNeill, Patricia D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.455C11  
CURRENT APPLICATION NUMBER: US/09/643,597  
CURRENT FILING DATE: 2000-08-21  
NUMBER OF SEQ ID NOS: 369  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 130  
LENGTH: 5156  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-643-597-130

Query Match 92.0%; Score 18.4; DB 4; Length 5156;  
Best Local Similarity 95.0%; Pred. No. 0.66;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGCGATGAGAA 20  
||||| |||||||  
DB 1356 TGTATTTCAGGCGATGAGAA 1375

RESULT 4  
US-08-317-450B-12  
Sequence 12, Application US/08317450B  
Patent No. 5660962  
GENERAL INFORMATION:  
APPLICANT: Trygvason, Karl  
APPLICANT: Kallunki, Pekka  
APPLICANT: Pyke, Charles  
TITLE OF INVENTION: Laminin Chains: Diagnostic and  
TITLE OF INVENTION: Therapeutic Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER & ALLEGRETTI, LTD.  
STREET: Ten South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/317,450B  
APPLICATION NUMBER: US/08/317,450B  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Chao, Mark  
REGISTRATION NUMBER: 37,293  
REFERENCE/DOCKET NUMBER: 94,778  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5200 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 118..183  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 118..3699  
FEATURE:  
NAME/KEY: polyA\_site  
LOCATION: 4433  
FEATURE:  
NAME/KEY: polyA\_site  
LOCATION: 5195  
US-08-317-450B-12

Query Match 92.0%; Score 18.4; DB 1; Length 5200;  
Best Local Similarity 95.0%; Pred. No. 0.66;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTTCAGGGGATGAGAA 20  
DB 1408 TGTTCAGGGGATGAGAA 1427

RESULT 5  
US-08-800-593-12  
Sequence 12, Application US/08800593  
Patent No. 6143505  
GENERAL INFORMATION:  
APPLICANT: Trygvason, Karl  
APPLICANT: Kallunki, Pekka  
APPLICANT: Pyke, Charles  
TITLE OF INVENTION: Laminin Chains: Diagnostic and  
TITLE OF INVENTION: Therapeutic Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff  
STREET: 300 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,593  
FILING DATE: 18-FEB-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/317,450  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Chao, Mark  
REGISTRATION NUMBER: 37,293  
REFERENCE/DOCKET NUMBER: 94,778-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-913-0001  
TELEFAX: 312-913-0002  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5200 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 118..183  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 118..3699  
FEATURE:  
NAME/KEY: polyA\_site  
LOCATION: 4433  
FEATURE:  
NAME/KEY: polyA\_site  
LOCATION: 5195  
US-08-800-593-12

Query Match 92.0%; Score 18.4; DB 3; Length 5200;  
Best Local Similarity 95.0%; Pred. No. 0.66;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTTCAGGGGATGAGAA 20  
DB 1408 TGTTCAGGGGATGAGAA 1427

RESULT 6  
US-09-381-150A-1/C  
Sequence 1, Application US/09381150A  
Patent No. 6465718  
GENERAL INFORMATION:  
APPLICANT: Inze, Dirk  
APPLICANT: Segers, Gerda  
APPLICANT: De Veylder, Lieven  
APPLICANT: Mironov, Vladimir  
TITLE OF INVENTION: METHOD AND MEANS FOR MODULATING  
CELL GROWTH AND CONTROL  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Nixon Peabody LLP  
STREET: 990 Stewart Avenue  
CITY: Garden City  
STATE: New York, New York  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/381,150A  
FILING DATE: 13-MAR-2000  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP PCT/EP98/01522  
FILING DATE: 13-MAR-1998  
APPLICATION NUMBER: EP 97.200.765.2  
FILING DATE: 14-MAR-1997  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 454 base pairs



ADDRESSEE: DYKEMA GOSSETT  
STREET: 400 Renaissance Center  
CITY: Detroit  
STATE: MI  
COUNTRY: USA  
ZIP: 48243  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,201  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/913,107  
FILING DATE: 14-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Kelly, Robert L.  
REGISTRATION NUMBER: 31,843  
REFERENCE/DOCKET NUMBER: 61,686-016  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (313) 540-0849  
TELEFAX: (313) 540-0763  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1810 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 177..1145  
US-08-459-201-1

Query Match 76.0%; Score 15.2; DB 1; Length 1810;  
Best Local Similarity 85.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTACTCAGGCGATGAGAA 20  
||||| ||||||| |||||  
Db 590 TGTTCCTAAGGGAAGAA 571

RESULT 10  
US-08-281-248-1/c  
Sequence 1, Application US/08281248  
Patent No. 5650500  
GENERAL INFORMATION:  
APPLICANT: Raz, Avraham  
APPLICANT: Nabl, Ivan R.  
APPLICANT: Otto, Thomas  
APPLICANT: Watanabe, Hideomi  
TITLE OF INVENTION: Method of Determining Metastatic  
TITLE OF INVENTION: Potential of Tumor Cells  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESS: DYKEMA GOSSETT  
STREET: 400 Renaissance Center  
CITY: Detroit  
STATE: MI  
COUNTRY: USA  
ZIP: 48243  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/281,248  
FILING DATE: 27-JUL-1994

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/913,107  
FILING DATE: 14-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Kelly, Robert L.  
REGISTRATION NUMBER: 31,843  
REFERENCE/DOCKET NUMBER: 61,686-016  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (313) 540-0849  
TELEFAX: (313) 540-0763  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1810 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 177..1145  
US-08-281-248-1

Query Match 76.0%; Score 15.2; DB 1; Length 1810;  
Best Local Similarity 85.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTACTCAGGCGATGAGAA 20  
||||| ||||||| |||||  
Db 590 TGTTCCTAAGGGAAGAA 571

RESULT 11  
US-09-090-793-13  
Sequence 13, Application US/09090793  
Patent No. 6140486  
GENERAL INFORMATION:  
APPLICANT: Calgene, LLC  
TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
TITLE OF INVENTION: of polyketide-like synthesis genes in plants  
FILE REFERENCE: CGNE.131.0105  
CURRENT FILING DATE: 1998-06-04  
EARLIER FILING DATE: 1997-06-04  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 13  
LENGTH: 19227  
TYPE: DNA  
ORGANISM: Vibrio marinus  
US-09-090-793-13

Query Match 76.0%; Score 15.2; DB 3; Length 19227;  
Best Local Similarity 85.0%; Pred. No. 47;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTACTCAGGCGATGAGAA 20  
||||| ||||||| |||||  
Db 15171 TGTACTCAGGCGATGAGAA 15190

RESULT 12  
US-09-090-793-12  
Sequence 12, Application US/09090793  
Patent No. 6140486  
GENERAL INFORMATION:  
APPLICANT: Calgene, LLC  
TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
TITLE OF INVENTION: of polyketide-like synthesis genes in plants  
FILE REFERENCE: CGNE.131.0105  
CURRENT FILING DATE: 1998-06-04  
APPLICATION NUMBER: US/09/090,793

EARLIER APPLICATION NUMBER: 60/048,650  
EARLIER FILING DATE: 1997-06-04  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 12  
LENGTH: 40138  
TYPE: DNA  
ORGANISM: Vibrio marinus  
US-09-090-793-12

Query Match 76.0%; Score 15.2; DB 3; Length 40138;  
Best local Similarity 85.0%; Pred. No. 34;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTACTCAGGAGATGAGAA 20  
||||| ||||| |||||  
Db 32154 TGTACTCAGGAGATGAGAA 32173

RESULT 13  
US-09-128-155-17/C  
Sequence 17, Application US/09128155  
Patent No. 6117654  
GENERAL INFORMATION:  
APPLICANT: Pan, Yang  
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
FILE REFERENCE: 09404/052001  
CURRENT APPLICATION NUMBER: US/09/128,155  
EARLIER FILING DATE: 1998-08-03  
EARLIER APPLICATION NUMBER: US 60/091,650  
EARLIER FILING DATE: 1998-07-02  
EARLIER APPLICATION NUMBER: US 60/054,646  
EARLIER FILING DATE: 1997-08-04  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 17  
LENGTH: 176373  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(176373)  
OTHER INFORMATION: n - A,T,C or G  
US-09-128-155-17

Query Match 76.0%; Score 15.2; DB 3; Length 176373;  
Best local Similarity 85.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTACTCAGGAGATGAGAA 20  
||||| ||||| |||||  
Db 7658 TTTGACTCAGGAGATGAGAA 7639

RESULT 14  
US-08-765-081-12  
Sequence 12, Application US/08765081  
Patent No. 5798260  
GENERAL INFORMATION:  
APPLICANT: Tarr, P.T., Bilge, S.S., Besser, T.E., Vary Jr., J.C.  
TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
STREET: 2800 Pacific First Center, 1420 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage  
COMPUTER: IBM PC compatible/Pentium

OPERATING SYSTEM: MS-WINDOW 3.1  
SOFTWARE: Word for Windows-6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/765,081  
FILING DATE: March 26, 1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/06994  
FILING DATE: June 7, 1995  
APPLICATION NUMBER: US 08/265,714  
FILING DATE: June 24, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Shelton, Dennis K.  
REGISTRATION NUMBER: 26,997  
REFERENCE/DOCKET NUMBER: CHOR-1-10286  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-206-682-8100; 1-206-224-0779 (direct)  
TELEFAX: 1-206-224-0779  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 300 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
DESCRIPTION: Genomic DNA fragment described on page 10  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli 0157:H7  
STRAIN: A5,F4,N11  
US-08-765-081-12

Query Match 74.0%; Score 14.8; DB 1; Length 300;  
Best local Similarity 88.9%; Pred. No. 36;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTACTCAGGAGATGAGAA 20  
||||| ||||| |||||  
Db 66 TAACACAGTGTGATGAGAA 83

RESULT 15  
US-09-098-082-12  
Sequence 12, Application US/09098082  
Patent No. 6040421  
GENERAL INFORMATION:  
APPLICANT: Tarr, P.T., Bilge, S.S., Besser, T.E., Vary Jr., J.C.  
TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
STREET: 2800 Pacific First Center, 1420 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage  
COMPUTER: IBM PC compatible/Pentium II  
OPERATING SYSTEM: MS-Windows 95  
SOFTWARE: Word for Windows-6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/098,082  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/765,081  
FILING DATE: March 26, 1997  
APPLICATION NUMBER: PCT/US95/06994  
FILING DATE: June 7, 1995  
APPLICATION NUMBER: US 08/265,714  
FILING DATE: June 24, 1994  
ATTORNEY/AGENT INFORMATION:







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 06:30:26 ; Search time 67.1795 seconds  
(without alignments)  
441.935 Million cell updates/sec

Title: US-10-053-662a-29  
Perfect score: 20  
Sequence: 1 tttactcaggagatgagaa 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCr\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCrUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	92.0	4316	9 US-10-227-738-14	Sequence 14, Appl
2	18.4	92.0	4316	9 US-09-756-071B-14	Sequence 14, Appl
3	18.4	92.0	5156	10 US-09-735-705-130	Sequence 130, App
4	18.4	92.0	5156	10 US-09-850-716A-130	Sequence 130, App
5	18.4	92.0	5156	10 US-09-897-778-130	Sequence 130, App
6	18.4	92.0	5175	9 US-10-171-311-114	Sequence 114, App
7	18.4	92.0	5200	9 US-10-227-738-12	Sequence 12, Appl
8	18.4	92.0	5200	10 US-09-756-071B-12	Sequence 12, Appl
9	18.4	92.0	5460	12 US-10-044-090-558	Sequence 558, App
10	16.8	84.0	186957	9 US-10-185-770-3	Sequence 3, Appl
11	16	80.0	22400	9 US-10-091-504-1385	Sequence 1385, Ap
12	16	80.0	22400	10 US-09-764-869-1385	Sequence 1385, Ap
13	15.8	79.0	12258	10 US-09-801-574-58	Sequence 58, Appl
14	15.4	77.0	486	10 US-09-783-590-4920	Sequence 4920, Ap
15	15.2	76.0	303	10 US-09-764-877-291	Sequence 291, App
16	15.2	76.0	425	10 US-09-764-877-2589	Sequence 2589, App
17	15.2	76.0	430	10 US-09-783-590-10809	Sequence 10809, A
18	15.2	76.0	454	9 US-10-225-966-1	Sequence 1, Appl
19	15.2	76.0	454	10 US-09-938-342-1	Sequence 1, Appl

C 20	15.2	76.0	489	9 US-09-918-995-28552	Sequence 28552, A
C 21	15.2	76.0	525	9 US-09-918-995-15287	Sequence 15287, A
C 22	15.2	76.0	696	9 US-10-225-966-2	Sequence 2, Appl
C 23	15.2	76.0	696	10 US-09-938-342-2	Sequence 2, Appl
C 24	15.2	76.0	700	10 US-09-770-149-239	Sequence 239, App
C 25	15.2	76.0	1208	9 US-10-012-542-14	Sequence 14, Appl
C 26	15.2	76.0	1810	9 US-10-153-666-239	Sequence 239, App
C 27	15.2	76.0	1962	9 US-10-062-831-24	Sequence 24, Appl
C 28	15.2	76.0	2335	10 US-09-917-860A-1569	Sequence 1569, Ap
C 29	15.2	76.0	2898	9 US-10-153-668-359	Sequence 359, App
C 30	15.2	76.0	3296	9 US-10-153-668-357	Sequence 357, App
C 31	15.2	76.0	3398	9 US-10-198-846-10944	Sequence 10944, A
C 32	15.2	76.0	3589	9 US-09-764-891-9611	Sequence 9611, Ap
C 33	15.2	76.0	6057	9 US-10-331-061-79	Sequence 79, Appl
C 34	15.2	76.0	19227	9 US-10-331-061-13	Sequence 13, Appl
C 35	15.2	76.0	21000	9 US-09-975-123-11	Sequence 11, Appl
C 36	15.2	76.0	40138	9 US-10-331-061-12	Sequence 12, Appl
C 37	15.2	76.0	78056	9 US-10-109-551-1	Sequence 1, Appl
C 38	15.2	76.0	176373	9 US-10-095-407-17	Sequence 17, Appl
C 39	15	75.0	206	10 US-09-864-761-4926	Sequence 4926, Ap
C 40	15	75.0	442	10 US-10-037-270-313	Sequence 313, App
C 41	15	75.0	1825	9 US-09-867-701-5313	Sequence 5313, Ap
C 42	14.8	74.0	439	10 US-09-918-995-6887	Sequence 6887, Ap
C 43	14.8	74.0	443	9 US-09-983-965-2432	Sequence 2432, Ap
C 44	14.8	74.0	451	10 US-09-983-965-2432	Sequence 450, App
C 45	14.8	74.0	502	10 US-09-783-590-450	

#### ALIGNMENTS

RESULT 1  
US-10-227-738-14  
Sequence 14, Application US/10227738  
Publication No. US20030100529A1  
GENERAL INFORMATION:  
APPLICANT: Tryggyrason, Karl  
Kallunki, Pekka  
Pyke, Charles  
TITLE OF INVENTION: Laminin Chains: Diagnostic and  
Therapeutic Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
STREET: 300 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/227,738  
FILING DATE: 26-Aug-2002  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,593  
FILING DATE: 18-FEB-1997  
APPLICATION NUMBER: US 08/317,450  
FILING DATE: 04-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Chao, Mark  
REGISTRATION NUMBER: 37,293  
REFERENCE/DOCKET NUMBER: 94,778-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-913-0001  
TELEFAX: 312-913-0002  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4316 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 118..183  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 118..3453  
FEATURE:  
NAME/KEY: repeat\_unit  
LOCATION: 4021..4316  
OTHER INFORMATION: /rpt\_type="other"  
/rpt\_family="HUMAN ALT"  
FEATURE:  
NAME/KEY: polyA\_site  
LOCATION: 4296  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-10-227-738-14

Query Match 92.0%; Score 18.4; DB 9; Length 4316;  
Best Local Similarity 95.0%; Pred. No. 3.8;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGGGATGAGAA 20  
||||| |||||||  
DB 1408 TGTATTTCAGGGGATGAGAA 1427

RESULT 2  
US-09-756-071B-14  
Sequence 14, Application US/09756071B  
Patent No. US20020052307A1  
- GENERAL INFORMATION:  
APPLICANT: Tryggyvason, Karl  
Kallunki, Pekka  
Pyke, Charles  
TITLE OF INVENTION: Laminin Chains: Diagnostic Uses  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fay Sharpe Fagan Minnich & McKee  
STREET: 1100 Superior Ave, Suite 700  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/756,071B  
FILING DATE: 08-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/663,147  
FILING DATE: 150-September 2000  
ATTORNEY/AGENT INFORMATION:  
NAME: Minnich, Richard, J  
REGISTRATION NUMBER: 24,175  
REFERENCE/DOCKET NUMBER: TRV 20014  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 216-861-5582  
TELEFAX: 216-241-1666  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4316 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 118..183  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 118..3453  
FEATURE:  
NAME/KEY: repeat\_unit  
LOCATION: 4021..4316  
OTHER INFORMATION: /rpt\_type="other"  
/rpt\_family="HUMAN ALT"  
FEATURE:  
NAME/KEY: polyA\_site  
LOCATION: 4296  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-09-756-071B-14

Query Match 92.0%; Score 18.4; DB 10; Length 4316;  
Best Local Similarity 95.0%; Pred. No. 3.8;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGGGATGAGAA 20  
||||| |||||||  
DB 1408 TGTATTTCAGGGGATGAGAA 1427

RESULT 3  
US-09-735-705-130  
Sequence 130, Application US/09735705  
Patent No. US20020052329A1  
- GENERAL INFORMATION:  
APPLICANT: Wang, Tonglong  
APPLICANT: Fan, Liqun  
APPLICANT: Kalos, Michael D.  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: Hosken, Nancy  
APPLICANT: Fanger, Gary R.  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Henderson, Robert A.  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Fanger, Neil  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.455C14  
CURRENT APPLICATION NUMBER: US/09/735,705  
CURRENT FILING DATE: 2000-12-12  
NUMBER OF SEQ ID NOS: 419  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 130  
LENGTH: 5156  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-735-705-130

Query Match 92.0%; Score 18.4; DB 10; Length 5156;  
Best Local Similarity 95.0%; Pred. No. 3.9;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGGGATGAGAA 20  
||||| |||||||  
DB 1356 TGTATTTCAGGGGATGAGAA 1375

RESULT 4  
US-09-850-716A-130  
Sequence 130, Application US/09850716A  
Patent No. US20020115139A1  
- GENERAL INFORMATION:  
APPLICANT: Kalos, Michael D.  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Retter, Marc W.

```

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C15
; CURRENT APPLICATION NUMBER: US/09/850,716A
; CURRENT FILING DATE: 2001-05-07
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 130
; LENGTH: 5156
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-850-716A-130

Query Match          92.0%; Score 18.4; DB 10; Length 5156;
Best Local Similarity 95.0%; Pred. No.3.9;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 TGTTCATCAGGGGATGAGAA 20
        ||||| ||||| ||||| |||||
Db      1356 TGTATTCAGGGGATGAGAA 1375

RESULT 5
US-09-897-778-130
; Sequence 130, Application US/09897778
; Patent No. US20020147143A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Peckham, David W.
; APPLICANT: Fanger, Neil
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C16
; CURRENT APPLICATION NUMBER: US/09/897,778
; CURRENT FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 467
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 130
; LENGTH: 5156
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-897-778-130

Query Match          92.0%; Score 18.4; DB 10; Length 5156;
Best Local Similarity 95.0%; Pred. No.3.9;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 TGTTCATCAGGGGATGAGAA 20
        ||||| ||||| ||||| |||||
Db      1356 TGTATTCAGGGGATGAGAA 1375

RESULT 6
US-10-171-311-114
; Sequence 114, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Ganavathapu, Manjula
; APPLICANT: Hoerssh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
```

```

; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114
; LENGTH: 5175
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-171-311-114

Query Match          92.0%; Score 18.4; DB 9; Length 5175;
Best Local Similarity 95.0%; Pred. No.3.9;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 TGTTCATCAGGGGATGAGAA 20
        ||||| ||||| ||||| |||||
Db      1380 TGTATTCAGGGGATGAGAA 1399

RESULT 7
US-10-227-738-12
; Sequence 12, Application US/10227738
; Publication No. US20030100529A1
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; Therapeutic Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/227,738
; FILING DATE: 26-Aug-2002
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,593
; FILING DATE: 18-FEB-1997
; APPLICATION NUMBER: US 08/317,450
; FILING DATE: 04-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 94,778-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
```

```
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3699
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4433
FEATURE:
NAME/KEY: polyA_site
LOCATION: 5195
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-227-738-12

Query Match
Best Local Similarity 92.0%; Score 18.4; DB 9; Length 5200;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20
Db 1408 TGTATTTCAGGGGATGAGAA 1427

RESULT 8
US-09-756-071B-12
Sequence 12, Application US/09756071B
Patent No. US20020052307A1
GENERAL INFORMATION:
APPLICANT: Trygvason, Karl
Kallunki, Pekka
Pyke, Charles
TITLE OF INVENTION: Laminin Chains: Diagnostic Uses
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ray Sharpe Fagan Minnich & McKee
STREET: 1100 Superior Ave, Suite 700
CITY: Cleveland
STATE: Ohio
COUNTRY: USA
ZIP: 44114
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/756,071B
FILING DATE: 08-Jan-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/663,147
FILING DATE: 150-September 2000
ATTORNEY/AGENT INFORMATION:
NAME: Minnich, Richard, J
REGISTRATION NUMBER: 24,175
REFERENCE/DOCKET NUMBER: TRV 20014
TELECOMMUNICATION INFORMATION:
TELEPHONE: 216-861-5582
TELEFAX: 216-241-1666
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 5200 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3699
```

```
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4433
FEATURE:
NAME/KEY: polyA_site
LOCATION: 5195
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-756-071B-12

Query Match
Best Local Similarity 92.0%; Score 18.4; DB 10; Length 5200;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20
Db 1408 TGTATTTCAGGGGATGAGAA 1427

RESULT 9
US-10-044-090-558
Sequence 558, Application US/10044090
Patent No. US20020137081A1
GENERAL INFORMATION:
APPLICANT: Olga Bandman
TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
FILE REFERENCE: PA-0028 US
CURRENT APPLICATION NUMBER: US/10/044,090
CURRENT FILING DATE: 2002-01-09
NUMBER OF SEQ ID NOS: 850
SOFTWARE: PERL Program
SEQ ID NO 558
LENGTH: 5460
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20020137081A1 149791.5
US-10-044-090-558

Query Match
Best Local Similarity 92.0%; Score 18.4; DB 12; Length 5460;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20
Db 1410 TGTATTTCAGGGGATGAGAA 1429

RESULT 10
US-10-185-770-3
Sequence 3, Application US/10185770
Publication No. US20030022217A1
GENERAL INFORMATION:
APPLICANT: CRECARDI, Toni et al.
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS, AND
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
FILE REFERENCE: CL0001247
CURRENT APPLICATION NUMBER: US/10/185,770
CURRENT FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: 60/301,852
PRIOR FILING DATE: 2001-07-02
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 186957
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(186957)
OTHER INFORMATION: n = A,T,C or G
US-10-185-770-3
```

Query Match 84.0%; Score 16.8; DB 9; Length 186957;  
Best Local Similarity 90.0%; Pred. No. 31;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20  
DB 98798 TTTTATTCAGGGGATGAGAA 98817

RESULT 11  
US-10-091-504-1385/C  
; Sequence 1385, Application US/10091504  
; Publication No. US20030059908A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC007C1  
; CURRENT APPLICATION NUMBER: US/10/091,504  
; CURRENT FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 2442  
; Prior Application removed - See File Wrapper or Palm  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1385  
; LENGTH: 22400  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-091-504-1385

Query Match 80.0%; Score 16; DB 9; Length 22400;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TTACTCAGGGGATGAG 18  
DB 944 TTACTCAGGGGATGAG 929

RESULT 12  
US-09-764-869-1385/C  
; Sequence 1385, Application US/09764869  
; Patent No. US20020061521A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC007  
; CURRENT APPLICATION NUMBER: US/09/764,869  
; CURRENT FILING DATE: 2001-01-17  
; Prior Application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2442  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1385  
; LENGTH: 22400  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-869-1385

Query Match 80.0%; Score 16; DB 10; Length 22400;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TTACTCAGGGGATGAG 18  
DB 944 TTACTCAGGGGATGAG 929

RESULT 13  
US-09-801-574-58  
; Sequence 58, Application US/09801574  
; Patent No. US20020081592A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Peijiang Jeremy  
; APPLICANT: Page, David C.

; TITLE OF INVENTION: Reproduction-Specific Genes  
; FILE REFERENCE: 0399, 2007-002  
; CURRENT APPLICATION NUMBER: US/09/801,574  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: 60/187,518  
; PRIOR FILING DATE: 2000-03-07  
; PRIOR APPLICATION NUMBER: 60/261,557  
; PRIOR FILING DATE: 2001-01-12  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 58  
; LENGTH: 12258  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-801-574-58

Query Match 79.0%; Score 15.8; DB 10; Length 12258;  
Best Local Similarity 89.5%; Pred. No. 92;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGA 19  
DB 9217 TGTACTCAGGGGATGAGA 9235

RESULT 14  
US-09-783-590-4920/C  
; Sequence 4920, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16, 2C1  
; CURRENT APPLICATION NUMBER: US/09/783,590  
; CURRENT FILING DATE: 2000-02-15  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1995-04-12  
; PRIOR APPLICATION NUMBER: 08/346,731  
; PRIOR FILING DATE: 1994-11-21  
; NUMBER OF SEQ ID NOS: 12485  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4920  
; LENGTH: 486  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (10)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (100)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (106)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (254)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (266)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (273)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (274)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (314)

```
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (315)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (320)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (321)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (329)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (331)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (349)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (350)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (353)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (354)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (356)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (358)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (359)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (366)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (369)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (370)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (380)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (381)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (392)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (396)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (401)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (403)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (406)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (410)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (417)
OTHER INFORMATION: n equals a,t,g, or c
```

```
NAME/KEY: misc feature
LOCATION: (424)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (425)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (427)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (430)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (432)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (442)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (443)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (444)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (449)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (450)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (455)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (457)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (459)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (474)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (475)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (477)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-4920
```

```
Query Match 77.0%; Score 15.4; DB 10; Length 486;
Best Local Similarity 94.1%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 4 TACTCAGGGATGAGAA 20
Db 243 TACTCTGGGATGAGAA 227
```

```
RESULT 15
US-09-764-877-291/c
Sequence 291, Application US/09764877
Patent No. US20020147140A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005
CURRENT APPLICATION NUMBER: US/09/764,877
PRIORITY FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 291
```



```

: LENGTH: 303
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (205)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (267)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (293)
: OTHER INFORMATION: n equals a,t,g, or c
: US-09-764-877-291

```

```

Query Match          76.0%; Score 15.2; DB 10; Length 303;
Best Local Similarity 85.0%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 TGTACTCAGGGGATGAGAA 20
         ||||| ||||| |||||
DB      83 TGTGCTGAGGGGTTGAGAA 64

```

Search completed: July 1, 2003, 07:26:17  
 Job time : 68.1795 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 03:57:39 ; Search time 904.615 Seconds  
(without alignments)  
358.063 Million cell updates/sec

Title: US-10-053-662a-29

Perfect score: 20

Sequence: 1 tgttaccagggagatgagaa 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rnd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	DB ID	Description
c 1	18.4	92.0	261 14	BQ325508 QV2-C1010
c 2	18.4	92.0	413 14	T70002 ycl19f04.r1
c 3	18.4	92.0	530 9	AT1710603 DKFZP6860
c 4	18.4	92.0	580 9	AU145040 AU145040
c 5	18.4	92.0	610 9	AL710604 DKFZP6860
c 6	18.4	92.0	782 12	BG679991 602626709

Result No.	Score	Query Match length	DB ID	Description
c 7	18.4	92.0	1007 12	BG749494 602707627
c 8	17.4	87.0	446 17	AO880436 HS_5044_A
c 9	17.4	87.0	560 17	AO481409 RPCT-11-2
c 10	16.8	84.0	194 17	A2328393 IM0052610
c 11	16.8	84.0	242 9	AA055712 z175e05.r
c 12	16.8	84.0	263 9	AV227530 AV227530
c 13	16.8	84.0	385 17	AQ264952 CITBI-E1-
c 14	16.8	84.0	437 17	AQ002138 CIT-HSP-2
c 15	16.8	84.0	442 10	BE247528 TCRAPIE3
c 16	16.8	84.0	512 17	A2169304 SP_0113_B
c 17	16.8	84.0	550 14	BQ320138 PM3-CT080
c 18	16.8	84.0	576 17	TA665812P T. brucei
c 19	16.8	84.0	628 17	A2992707 2M0277022
c 20	16.8	84.0	822 12	BF300476 BF300476
c 21	16.8	84.0	893 14	BQ420274 AGENCOURT
c 22	16.8	84.0	914 12	BF671760 BF671760
c 23	16.8	84.0	1044 12	BE748740 601571905
c 24	16.6	83.0	958 17	CNS02A60 CNS02A60
c 25	16.4	82.0	60 17	CGA350111 CGA350111
c 26	16.4	82.0	269 17	AO103499 HS_3058_B
c 27	16.4	82.0	382 17	BH364764 CH230-60P
c 28	16.4	82.0	385 17	A2734473 RPCT-24-1
c 29	16.4	82.0	398 12	BG100888 uy16f08.y
c 30	16.4	82.0	840 17	BH554244 BOGEE83TF
c 31	16.4	82.0	342 12	BG043081 stf2a02.y
c 32	16.4	82.0	468 10	BE554551 ur48c05.x
c 33	16.4	82.0	752 17	BH384164 AG-ND-156
c 34	15.8	79.0	172 9	AV093257 AV093257
c 35	15.8	79.0	182 13	BH091644 saq99f12.
c 36	15.8	79.0	210 17	AO310289 CITBI-E1-
c 37	15.8	79.0	229 14	F05097 HSC01C011
c 38	15.8	79.0	263 17	A2942797 2M0203009
c 39	15.8	79.0	301 14	B58268 C0908 Fetal
c 40	15.8	79.0	348 10	BE465362 hm15C03.x
c 41	15.8	79.0	351 17	B96116 F23P3TR IG
c 42	15.8	79.0	378 14	W03415 za07e12.r1
c 43	15.8	79.0	390 17	AO060420 CIT-HSP-2
c 44	15.8	79.0	410 14	D39211 R1CR3041A R
c 45	15.8	79.0	411 17	AQ305163 HS_2022_A

## ALIGNMENTS

RESULT 1  
BQ325508/c  
LOCUS  
QV2-C10104-020301-611-d06 C10104 Homo sapiens CDNA, mRNA sequence.  
ACCESSION  
BQ325508.1 GI:20939888  
VERSION  
EST.  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 261)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=GV26f2-GV2-C10104-020301-611-d066f3-2001-03-02&f4-1)

Seq primer: puc 18 forward  
High quality sequence start: 5  
High quality sequence stop: 261.  
Location/Qualifiers

## FEATURES

source

1. 261  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="C10104"  
/dev\_stage="Adult"  
/note="Organ: colon\_lns; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESSES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

## BASE COUNT

54 a 79 c 76 g 52 t

## ORIGIN

Query Match 92.0%; Score 18.4; DB 14; Length 261;  
Best Local Similarity 95.0%; Pred. No. 76;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## QY

1 TGTACTCAGGCGATGACAA 20  
||||| |||||||

Db 211 TGTATTACGGCGATGACAA 192

## RESULT 2

LOCUS T70002 413 bp mRNA linear EST 23-FEB-1995  
DEFINITION ycl9f04.r1 Stratagene lung (#937210) Homo sapiens cDNA clone  
IMAGE:81151 5' similar to contains MER20 repetitive element ;, mRNA  
sequence.  
T70002  
T70002.1 GI:681150

## VERSION

EST.

## KEYWORDS

## SOURCE

## ORGANISM

human.

## REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Site: 936

High quality sequence stops: 331 Source: IMAGE Consortium, LNL. This

clone is available royalty-free through LNL; contact the IMAGE

Consortium (info@image.llnl.gov) for further information.

Insert length: 936 Std Error: 0.00

Seq primer: M13Rpl

High quality sequence stop: 331.

Location/Qualifiers

1. 413

/organism="Homo sapiens"

/db\_xref="GDB:484768"

/db\_xref="taxon:9606"  
/clone\_image:81151"  
/clone\_lib="Stratagene lung (#937210)"  
/sex="male"  
/dev\_stage="72 years"

/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: lung; Vector: pBluescript SK-; Site\_1: EcoRI  
; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
dT, normal lung. Average insert size: 1.0 kb; Uni-ZAP XR  
Vector: -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3'  
adaptor sequence: 5' CTCAGCTTTTCTTTTCTTTT 3'"

## BASE COUNT

121 a 83 c 104 g 102 t 3 others

## ORIGIN

Query Match 92.0%; Score 18.4; DB 14; Length 413;  
Best Local Similarity 95.0%; Pred. No. 89;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## QY

1 TGTACTCAGGCGATGACAA 20  
||||| |||||||

Db 230 TGTACTTAGGCGATGACAA 249

## RESULT 3

LOCUS AL710603 530 bp mRNA linear EST 22-MAR-2002  
DEFINITION DKFZP6860036.r1 686 (synonym: hlcc3) Homo sapiens cDNA clone  
DKFZP6860036 5', mRNA sequence.  
AL710603  
AL710603.1 GI:19693958

## VERSION

EST.

## KEYWORDS

## SOURCE

## ORGANISM

human.

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

MIPS

Am Klopferpitz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

Resequenced by EMBL (European Molecular Biology Laboratories),

Heidelberg/Germany) within the cDNA sequencing consortium of the

German Genome Project.

No sl sequence available.

This clone (DKFZP6860036) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

## FEATURES

source

1. 530

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="DKFZP6860036"

/clone\_lib="686 (synonym: hlcc3)"

/tissue\_type="human skeletal muscle"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Vector: pTriplEx2; Site\_1: SfiI; Site\_2: SfiIb;  
cDNA-collection"

BASE COUNT 168 a 109 c 145 g 108 t

## ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 530;

Best Local Similarity 95.0%; Pred. No. 97;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGCGATGACAA 20

||||| |||||||

Db 61 TGTACTTAGGCGATGACAA 80

RESULT 4  
LOCUS AUI45040/c 580 bp mRNA linear EST 05-AUG-2002  
DEFINITION AUI45040 HEMBAI Homo sapiens CDNA clone HEMBAI003720 3', mRNA sequence.  
ACCESSION AUI45040  
VERSION AUI45040.1 GI:11006561  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Salto,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.).  
HRI human CDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Salto,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano,S., Isogai,T.).  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human CDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; CDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.  
FEATURES  
source  
1..580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEMBAI003720"  
/clone\_lib="HEMBAI"  
/tissue\_type="whole embryo, mainly head"  
/dev\_stage="embryo, 10 weeks"  
/note="vector: pME18SFL3"  
BASE COUNT 122 a 156 c 126 g 175 t 1 others  
ORIGIN  
Query Match 92.0%; Score 18.4; DB 9; Length 580;  
Best Local Similarity 95.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 TGTACTCAGGGGATGAGAA 20  
||||| |||||||||  
DB 524 TGTACTTAGGGGATGAGAA 505  
RESULT 5  
LOCUS AL710604 610 bp mRNA linear EST 22-MAR-2002  
DEFINITION DKFZP6860046.L1 686 (synonym: hlc3) Homo sapiens CDNA clone DKFZP6860046 5', mRNA sequence.  
ACCESSION AL710604  
VERSION AL710604.1 GI:19693959  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Ansoorge,W., Winkler,U., Mewes,W., Well,B. and Wiemann,S.  
TITLE EST (Ansoorge,W., Winkler,U., Mewes,W., Well,B. and Wiemann,S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Ansoorge W  
MIPS  
Am Klopferspitz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
Sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the CDNA sequencing consortium of the German Genome Project.  
No sl sequence available.  
This clone (DKFZP6860046) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
FEATURES  
source  
1..610  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZP6860046"  
/clone\_lib="686 (synonym: hlc3)"  
/tissue\_type="human skeletal muscle"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="vector: pTR1p1Ex2; Site\_1: SfilA; Site\_2: SfilB; CDNA-collection"  
BASE COUNT 199 a 126 c 161 g 124 t  
ORIGIN  
Query Match 92.0%; Score 18.4; DB 9; Length 610;  
Best Local Similarity 95.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 TGTACTCAGGGGATGAGAA 20  
||||| |||||||||  
DB 64 TGTACTTAGGGGATGAGAA 83  
RESULT 6  
LOCUS BG679991 782 bp mRNA linear EST 01-MAY-2001  
DEFINITION 602626709F1 NCI\_CGAP\_Skn4 Homo sapiens CDNA clone IMAGE:4751499 5', mRNA sequence.  
ACCESSION BG679991  
VERSION BG679991.1 GI:13911388  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE NIH-MGC  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: James Cleaver, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: ILAM10608 row: a column: 04  
High quality sequence start: 6  
High quality sequence stop: 781.  
FEATURES  
source  
1..782  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4751499"  
/clone\_lib="NCI CGAP\_Skn4"  
/tissue\_type="squamous cell carcinoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPOK6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 179 a 194 c 231 g 178 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 12; Length 782;  
Best Local Similarity 95.0%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGAGATGAGAA 20  
|||||  
DB 127 TGTATTCAGGAGATGAGAA 146

RESULT 7  
LOCUS BG749494 1007 bp mRNA linear EST 15-MAY-2001  
DEFINITION 60207627f1 NTH\_MGC\_43 Homo sapiens cDNA clone IMAGE:484167 5',  
ACCESSION BG749494  
VERSION BG749494.1 GI:14060147  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1007)  
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LNCM1681 row: f column: 08  
High quality sequence stop: 714.  
Location/Qualifiers  
1..1007  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:484167"  
/clone\_lib="NTH\_MGC\_43"  
/tissue\_type="normal pigmented retinal epithelium"  
/lab\_host="DH10B (Phage-resistant)"  
/note="Organ: eye; Vector: pORF7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-CDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NTH\_MGC library."\*

BASE COUNT 232 a 270 c 309 g 196 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 12; Length 1007;  
Best Local Similarity 95.0%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGAGATGAGAA 20  
|||||  
DB 140 TGTATTCAGGAGATGAGAA 159

RESULT 8  
LOCUS AQ880436 446 bp DNA linear GSS 09-NOV-1999  
DEFINITION HS.5044.A2.D10.SP6E.RPCI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate-8812 Col-20 Row-G, DNA sequence.  
ACCESSION AQ880436

VERSION AQ880436.1 GI:6311903  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 446)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: [jwallace@u.washington.edu](mailto:jwallace@u.washington.edu)  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
([piet@edejong.med.buffalo.edu](mailto:piet@edejong.med.buffalo.edu)). Clones may be purchased from  
BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm))  
or from Resear h Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end web Server:  
<http://www.htsc.washington.edu>  
Plate: 8812 row: G column: 20  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 446.  
Location/Qualifiers  
1..446  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-8812 Col-20 Row-G"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI methylase. Size selected DNA was cloned into the  
pBACe3.6 vector at EcoRI sites"

BASE COUNT 116 a 101 c 68 g 159 t 2 others  
ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 446;  
Best Local Similarity 94.7%; Pred. No. 2.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GTTACTCAGGAGATGAGAA 20  
|||||  
DB 362 GCTACTCAGGAGATGAGAA 344

RESULT 9  
LOCUS AQ481409 560 bp DNA linear GSS 24-APR-1999  
DEFINITION RPCI-11-241H7.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-241H7,  
DNA sequence.  
ACCESSION AQ481409  
VERSION AQ481409.1 GI:4668813  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 560)  
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter  
J.C.  
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building (1997)

## COMMENT

Other GSSs: RPCI-11-241H7.TV  
 Contact: Shaying Zhao, William Niernman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: SP6  
 Class: BAC ends.

## FEATURES

Source Location/Qualifiers

1..560  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7592334"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-241H7"  
 /clone\_11b="RPCI-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"  
 BASE COUNT 143 a 128 c 88 g 201 t  
 ORIGIN

## Query Match

Best Local Similarity 94.7%; Score 17.4; DB 17; Length 560;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCTACTCAGGAGATGAGAA 20  
 1 |||||

Db 355 GCTACTCAGGAGATGAGAA 337

## RESULT 10

## LOCUS

AZ328393 194 bp DNA linear GSS 29-SEP-2000

DEFINITION 1M0052610F Mouse 10kb plasmid UUC1M library Mus musculus genomic

clone UUC1M0052610 F, DNA sequence.

ACCESSION AZ328393.1 GI:10388074

VERSION GSS.

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Mus musculus

REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS 1 (bases 1 to 194)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.  
 and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

CONTACT: Robert B. Weiss

UNIVERSITY OF UTAH GENOME CENTER

RM. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

TEL: 801 585 5606

FAX: 801 585 7177

EMAIL: ddunn@genetics.utah.edu

INSERT LENGTH: 10000 Std Error: 0.00

PLATE: 0052 row: G column: 10

SEQ PRIMER: CCTGTAAACAGACGCCACGT

CLASS: plasmid ends

HIGH QUALITY SEQUENCE STOP: 194.

LOCATION/QUALIFIERS

## source

1..194  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUC1M0052610"  
 /clone\_11b="Mouse 10kb plasmid UUC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: pMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g147321419b1AF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 56 a 32 c 44 g 62 t  
 ORIGIN

## Query Match

Best Local Similarity 90.0%; Score 16.8; DB 17; Length 194;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTACTCAGGAGATGAGAA 20  
 1 |||||

Db 18 TGTCTCTCAGGAGATGAGAA 37

## RESULT 11

## LOCUS

AA055712/c 242 bp mRNA linear EST 19-MAY-1997

DEFINITION 2175e05.r1 Stratagene colon (#337204) Homo sapiens CDNA clone

IMAGE:510464 5' similar to gb:s85655 PROHIBITIN (HUMAN);, mRNA

sequence.

ACCESSION AA055712.1 GI:1548050

VERSION AA055712.1

KEYWORDS EST.

SOURCE human

ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 242)  
 Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapell,B.,  
 Chisoe,S., Dietrich,N., Dubugue,T., Favell,A., Gish,W., Hawkins,  
 M., Hultman,M., Kucaba,T., Lacy,M., Le,N., Mardis,E., Moore,  
 B., Morris,M., Parsons,J., Prange,C., Rikkin,L., Rohlfing,T.,  
 Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,  
 Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Warr,M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

97044478

CONTACT: Willson RK

WASHINGTON UNIVERSITY SCHOOL OF MEDICINE

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

TEL: 314 286 1800

FAX: 314 286 1810

EMAIL: est@watson.wustl.edu

THIS CLONE IS AVAILABLE ROYALTY-FREE THROUGH LNL; CONTACT THE

IMAGE CONSORTIUM (info@image.lnl.gov) FOR FURTHER INFORMATION.

INSERT LENGTH: 1946 Std Error: 0.00

SEQ PRIMER: -28M13 rev2 from Amersham

HIGH QUALITY SEQUENCE STOP: 168.

FEATURES		location/Qualifiers	
SOURCE		1. .242	
		/organism="Homo sapiens"	
		/db_xref="GDB:381379"	
		/db_xref="taxon:9606"	
		/clone_1="IMAGE:510464"	
		/clone_1id="Stratagene colon (#937204)"	
		/tissue_type="tumor"	
		/cell_line="r84 carcinoma cell line"	
		/lab_host="SOBR cells (kanamycin resistant)"	
		/note="organ: colon; Vector: pBluescript SK+; Site:1: EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo dr. T-84 colonic epithelial cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"	
BASE COUNT	55 a	46 c	74 g 64 t 3 others
ORIGIN			
Query Match	84.0%;	Score 16.8;	DB 9; Length 242;
Best Local Similarity	90.0%;	Pred. No. 4.7e+02;	
Matches	18; Conservative	0; Mismatches	2; Indels 0; Gaps 0;
Oy	1 TGGTACCTGAGGGGATGAGAA 20		
Db			
	220 TGTTCACCGACGATGAGAA 201		
RESULT 12			
LOCUS	AV227530	263 bp	mRNA linear EST 14-NOV-2001
DEFINITION	AV227530 RIKEN full-length cDNA clone, 14 days embryo liver Mus		
VERSION	AV227530		
KEYWORDS	AV227530.1 GI:6179197		
SOURCE	EST.		
ORGANISM	house mouse.		
	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 263)		
AUTHORS	Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, M., Tomioka, N., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamuta, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.		
TITLE	RIKEN Mouse ESTs (Kono, H., et al. 1999)		
COMMENT	Unpublished (1999)		
COMMENT	Contact: Yoshihide Hayashizaki		
COMMENT	Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute		
COMMENT	The Institute of Physical and Chemical Research (RIKEN)		
COMMENT	1-7-22 Stehro-cho, Tsunumi-Ku, Yokohama, Kanagawa 230-0045, Japan		
COMMENT	Tel: 81-45-503-9222		
COMMENT	Fax: 81-45-503-9216		
COMMENT	Email: genome-res@sc.riken.go.jp,		
COMMENT	URL: http://genome.gsc.riken.go.jp/		
COMMENT	Sasaki, N., Izawa, M., Watabiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.		
COMMENT	Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)		
COMMENT	Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.		
COMMENT	Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)		
COMMENT	Carninci, P. and Hayashizaki, Y.		

<b>FEATURES</b>					
source	Location/Qualifiers 1..263 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="AA32406C24" /clone_lib="RIKEN full-length enriched, 14 days embryo liver" /liver /sex="mixed" /tissue_type="liver" /dev_stage="14 days embryo" /lab_host="DH10B" /note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15' GAGGAGAAGACGCATCCACAGCGCTTTTGTGGTTTTTTVVN 3'), cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGAGAGATTTCGAGTAATTAATAATGCCCCCCCCCCCCC 3'). cDNA was cloned into the XhoI and BamHI sites. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FIC I"				
BASE COUNT	65 a	60 c	50 g	88 t	
ORIGIN					
Query Match	84.0%; Score 16.8;	DB 9;	Length 263;		
Best Local Similarity	90.0%;	Pred. No. 4.8e+02;			
Matches 18;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;	
OY	1 TGTTACTCAGGGATGAGAA 20				
Dd	141 TGTACTGAGGGATGGGA 160				
<b>RESULT 13</b>					
A0264952/c	385 bp DNA linear GSS 27-Oct-1998				
LOCUS	CITBI-EI-2510J9.TF CITBI-EI Homo sapiens genomic clone 2510J9, DNA				
DEFINITION	sequence.				
ACCESSION	A0264952				
VERSION	A0264952.1 GI:3793152				
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 385) Adams,M.D., Rounsley,S.D., Zhao,S., Baas,S., Linher,K., Golden,K., Berry,K.K., Granger,D., Sun,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C. Use of a random human BAC End Sequence Database for Sequence-Ready Map Building Unpublished (1998)				
TITLE	Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: madams@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC				
JOURNAL COMMENT	Other-GSS: CITBI-EI-2510J9.TR				



end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).

Seq primer: M13-21  
Class: BAC ends.

**Class: BAC ends.**

FEATURES	Location/Qualifiers
source	1. .385

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="251079"
/clone_1lb="CITR1-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pReloBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltex Human PKC Library D"
BASE COUNT
103 a 117 c 57 g 108 t

```

RESULT	14
A0002138	
LOCUS	
DEFINITION	A0002138 437 bp DNA linear GSS-26-TUN-1998
CIT-HSP	2282A8.TR CIT-HSP Homo sapiens genomic clone 2282A8, DNA
sequence.	
A0002138	
VERSION	A0002138.1 GI:3029342
KEYWORDS	GSS.
SOURCE	human.

```

FEATURES
source
location/Qualifiers
1..437
/organism="Homo sapiens"
/db_xref="GDB:7147427"
/db_xref="taxon:9606"
/clone="3282A8"
/clone_11b="CTT-HSP"
/sex="Male"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelOBAc11; site_1: HindIII; site_2: HindIII"

```

Matches	18;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

1 TGTACTCAGGGATGAGAA 20

Db 204 TGTACTCAGGTGATAAGAA 223

RESULT 15					
BE247528/c					
LOCUS	BE247528	442 bp	mRNA	linear	EST-03-OCT-2001
DEFINITION	BCAP2E3364 Pediatric pre-B cell acute lymphoblastic leukemia TCBAP-E3364 project-TCBA Homo sapiens CDNA clone TCBPAP3364, mRNA sequence.				
ACCESSION	BE247528				
VERSION	BE247528..1	GI:9099285			
KEYWORDS	EST,				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eucharyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 442)				
AUTHORS	Weil,Y., Tsang,Y.T.M., Mel,G., Ku,J.M., Ali-Osman Jr.,F.R., Muzny,D., Bouck,J., Gibbs,R.A. and Margolin,J.F.				
TITLE	Pediatric Leukemia cDNA Sequencing Project				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Dr. Judith F. Margolin				

FEATURES	Location/Qualifiers
SOURCE	1. .442

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TCBAP3364"
/clone_lib="Pediatric pre-B cell acute lymphoblastic
leukemia Baylor-HESC project-TCBA"
/sex="male"
/tissue_type="Leukopheresis"
/cell_type="pre-B cell"
/dev_stage="pediatric 2 years"
/lab_host="DH10B"
/node="Vector: lambda psb; site_1: BamHI; site_2: EcoRI;
first strand cDNA was primed with an anchored
XhoI-oligo(dT) primer [5'GGAGGCTGAGGAGGGGGGAGGAGGAG(T)VN
3'; V-A,C,G; N-A,C,G,T] and then de tailed. Second strand
was primed with a BamHI-dC primer
[5'AGAGGCTGAGATCCGCGCGCCGATATATATATAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoI
and directionally cloned into the BamHI and SalI sites of
lambda psb vector. Library was then digested with BamHI and XhoI
at RIKEN
normalization. Library was constructed by Wei Yu at RIKEN
of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,
Itoh M, Nagaoka S, Sasakini, Okazaki Y, Muramatsu M,
Schneider C, Hayashizaki Y, High efficiency selection of
full-length cDNA by Improved biotinylated cap trapper.,
DNA Res 4: 1, 61-6, Feb 28, 1997)"

```

Query Match	84.08;	Score 16.8;	DB 17;	Length 4377
Best Local Similarity	90.08;	Pred. No. 5.8e+02;		

Tue Jul 1 11:47:46 2003

us-10-053-662a-29.rst

Page 8

Search completed: July 1, 2003, 07:03:41  
Job time : 909.615 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 05:18:35 ; Search time 556.359 Seconds  
(without alignments)  
993.879 Million cell updates/sec

Title: US-10-053-662A-30

Perfect score: 19

Sequence: 1 ctg999gcagttatgcac 19

Scoring table: IDENTITY NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

GenBank1:\*

1: gb\_ba:\*

2: gb\_hg:\*

3: gb\_in:\*

4: gb\_lm:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_hgtg\_hum:\*

31: em\_hgtg\_inv:\*

32: em\_hgtg\_other:\*

33: em\_hgtg\_mus:\*

34: em\_hgtg\_pln:\*

35: em\_hgtg\_rtd:\*

36: em\_hgtg\_mam:\*

37: em\_hgtg\_vrt:\*

38: em\_sy:\*

39: em\_hgtgo\_hum:\*

40: em\_hgtgo\_mus:\*

41: em\_hgtgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.4	91.6	67605	2	AC096816
C 2	17.4	91.6	100418	2	HS1028E2
C 3	17.4	91.6	102165	2	AC083819
C 4	17.4	91.6	158149	2	AC069000
C 5	17.4	91.6	163218	2	AL445989
C 6	17.4	91.6	171012	2	AC016061
C 7	17.4	91.6	202509	2	AC129875
C 8	17	89.5	168435	2	AC098629
C 9	17	89.5	170238	2	AC117955
C 10	16.4	86.3	2785	2	AK091176
C 11	16.4	86.3	7752	2	AB058743
C 12	16.4	86.3	30102	3	AF000195
C 13	16.4	86.3	38525	3	AF003145
C 14	16.4	86.3	42325	1	U00015
C 15	16.4	86.3	65707	2	AC101474
C 16	16.4	86.3	78576	2	AC127990
C 17	16.4	86.3	93307	2	AC125364
C 18	16.4	86.3	107889	2	AC004519
C 19	16.4	86.3	128484	9	AC025270
C 20	16.4	86.3	160394	2	AC012273
C 21	16.4	86.3	162495	2	AC009996
C 22	16.4	86.3	165423	2	AC120897
C 23	16.4	86.3	166975	9	AC069384
C 24	16.4	86.3	175465	2	AC009821
C 25	16.4	86.3	206644	2	AL732501
C 26	16.4	86.3	344050	1	MEPRN2
C 27	16	84.2	80398	9	AL513485
C 28	16	84.2	154101	2	AL537565
C 29	16	84.2	163284	9	AL137000
C 30	16	84.2	179024	2	AC112551
C 31	16	84.2	201254	2	AL626767
C 32	16	84.2	341560	2	AL596304
C 33	15.8	83.2	424	9	AF424831
C 34	15.8	83.2	1119	8	AF247165
C 35	15.8	83.2	1138	5	AF370640
C 36	15.8	83.2	1490	6	AX011679
C 37	15.8	83.2	2529	9	AF152487
C 38	15.8	83.2	2592	9	AB033004
C 39	15.8	83.2	2733	8	SCYLR12IC
C 40	15.8	83.2	4020	9	HSALJ7610
C 41	15.8	83.2	5260	9	AF152317
C 42	15.8	83.2	6387	9	AB0002343
C 43	15.8	83.2	6868	9	HSMB003621
C 44	15.8	83.2	23223	8	YSCLE9233
C 45	15.8	83.2	31694	9	AL353606

#### ALIGNMENTS

RESULT 1  
AC096816/c 67605 bp DNA linear HTG 11-JUL-2002  
LOCUS Rattus norvegicus clone CH230-17E18, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\* 40 nonordered pieces.  
ACCESSION AC096816 GI:21723712  
VERSION AC096816.3  
KEYWORDS HTG; HTGS; PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 67605)  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Ayele, M., Banks, T.,  
 Barbieri, J., Benton, J., Bimage, K., Blankenburg, K., Bonini, D.,  
 Bouche, J., Bowles, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
 Buhay, C., Burch, P., Burkett, C., Butrell, K.L., Byrd, N.C.,  
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gottlieb, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
 Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolyet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louissege, H.,  
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Meindale, A., Martinez, E.,  
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, K., Morris, S.,  
 Moser, M., Neal, D., Newton, E., Newkirk, S., Ogih, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, A., Nguyen, N.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pul, L., Quiles, M., Ren, Y.,  
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G.,  
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, J.,  
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, O.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, T., Wu, T.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Welstock, G., and Gibbs, R.

Unpublished  
 Direct Submission  
 2 (bases 1 to 67605)  
 Worley, K.C.

Direct Submission  
 Submitted (28-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 67605)

Direct Submission  
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 10, 2002 this sequence version replaced g1:17944640.

----- Genome Center

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GRVY  
 Center clone name: CH230-171E18

----- Summary Statistics

Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 26667 bases at least Q40  
 Consensus quality: 30535 bases at least Q30  
 Consensus quality: 33391 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently

\* consists of 40 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 1007 contig of 1006 bp in length  
 1107 gap of unknown length  
 1107 contig of 1484 bp in length  
 2590 gap of unknown length  
 2591 gap of unknown length  
 3591 contig of 1251 bp in length  
 3942 gap of unknown length  
 4042 contig of 1175 bp in length  
 5216 gap of unknown length  
 5317 contig of 1463 bp in length  
 6779 gap of unknown length  
 6780 gap of unknown length  
 7907 contig of 1028 bp in length  
 7908 gap of unknown length  
 8007 gap of unknown length  
 8008 contig of 1017 bp in length  
 9025 gap of unknown length  
 9125 contig of 1342 bp in length  
 10467 gap of unknown length  
 10467 contig of 1255 bp in length  
 11822 gap of unknown length  
 11822 contig of 1323 bp in length  
 13244 gap of unknown length  
 13345 gap of unknown length  
 15223 contig of 1879 bp in length  
 15324 gap of unknown length  
 15324 contig of 1500 bp in length  
 16823 gap of unknown length  
 16824 contig of 1078 bp in length  
 16924 gap of unknown length  
 18001 contig of 1278 bp in length  
 18102 gap of unknown length  
 19379 contig of 1278 bp in length  
 19479 gap of unknown length  
 19480 contig of 1893 bp in length  
 21372 gap of unknown length  
 21373 contig of 1414 bp in length  
 22887 gap of unknown length  
 22887 contig of 1583 bp in length  
 24570 gap of unknown length  
 24570 contig of 2079 bp in length  
 26749 gap of unknown length  
 26848 contig of 1254 bp in length  
 26849 gap of unknown length  
 28102 contig of 1505 bp in length  
 28203 gap of unknown length  
 29707 contig of 1505 bp in length  
 29708 gap of unknown length  
 30900 contig of 1093 bp in length  
 30901 gap of unknown length  
 31001 contig of 1406 bp in length  
 32407 gap of unknown length  
 32506 gap of unknown length  
 34962 contig of 2456 bp in length  
 34963 gap of unknown length  
 35062 gap of unknown length  
 36839 contig of 1777 bp in length  
 36840 gap of unknown length  
 36940 contig of 1788 bp in length  
 38727 gap of unknown length  
 38728 contig of 1496 bp in length  
 40323 gap of unknown length  
 40423 contig of 1429 bp in length  
 41853 gap of unknown length  
 41952 contig of 1820 bp in length  
 43772 gap of unknown length  
 43773 contig of 1584 bp in length  
 45456 gap of unknown length  
 45457 gap of unknown length  
 45557 contig of 1997 bp in length  
 47554 gap of unknown length  
 47554 contig of 1252 bp in length  
 47654 gap of unknown length  
 48905 contig of 1570 bp in length  
 49006 gap of unknown length  
 50575 contig of 1222 bp in length  
 50676 gap of unknown length  
 51897 contig of 1222 bp in length  
 51897 gap of unknown length

\* 51998 54197: contig of 2200 bp in length  
\* 54198 54297: gap of unknown length  
\* 54298 55860: contig of 1563 bp in length  
\* 55861 55960: gap of unknown length  
\* 55961 58193: contig of 2233 bp in length  
\* 58194 58293: gap of unknown length  
\* 58294 58961: contig of 1568 bp in length  
\* 58962 59961: gap of unknown length  
\* 59962 62224: contig of 2263 bp in length  
\* 62225 62324: gap of unknown length  
\* 62325 64813: contig of 2489 bp in length  
\* 64814 67605: gap of unknown length  
\* 64914 67605: contig of 2692 bp in length.  
Location/Qualifiers  
1. 67605  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-171E18"  
BASE COUNT 15603 a 14128 c 14795 g 14669 t 8410 others  
ORIGIN  
Query Match 91.6%; Score 17.4; DB 2; Length 67605;  
Best Local Similarity 94.7%; Pred. No. 70;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 CTGGGGGCGAGTTATGAC 19  
DB 24769 CTGGGGGCGAGTTATGAC 24751  
|||||  
RESULT 2  
HS1026E2 100418 bp DNA linear PRI 23-NOV-1999  
LOCUS HS1026E2/c  
DEFINITION Human DNA sequence from clone 1026E2 on chromosome 1q24.1-25.3 EST,  
CA repeat, STS, GSS, complete sequence.  
ACCESSION AL022143  
VERSION AL022143.1 GI:3319672  
KEYWORDS HTG; repeat polymorphism.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 100418)  
W1kinson/J.  
REFERENCE Direct Submission  
AUTHORS Submitted (08-JUL-1998) E-mail enquiries: humquerry@sanger.ac.uk  
TITLE Cloned request: clonerequest@sanger.ac.uk  
JOURNAL On Jul 15, 1998 this sequence version replaced gi:2969924.  
COMMENT During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences  
with only a small overlap as described above. IMPORTANT: This  
sequence is not the entire insert of clone 1026E2. It may be  
shorter because we only sequence overlapping sections once, or  
longer because we arrange for a small overlap between neighbouring  
submissions.  
The true left end of clone 964D12 is at 100315 in this sequence.  
The true right end of clone 35C21 is at 59922 in this sequence.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
This sequence was generated from part of bacterial clone contigs of  
human chromosome 1, constructed by the Sanger Centre Chromosome 1  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Ch1  
1026E2 is from the library RPI5 constructed at the Roswell Park  
Cancer Institute by the group of Pieter de Jong. For further  
details see http://bacpac.med.buffalo.edu/VECTOR: pcypac2.

FEATURES  
source  
Location/Qualifiers  
1. 100418  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="q24.1-25.3"  
/clone="RP5-1026E2"  
/clone\_1lb="RPI5"  
9. 1897  
/note="Retroviral Sequence"  
2018. 2177  
/note="2 copies 80 mer 95% conserved"  
2052. >2289  
/note="match: EST AA601759 clone IMAGE:1099629"  
2176. 2339  
/note="2 copies 82 mer 85% conserved"  
prim\_transcript complement(2623. 3021)  
/note="match: 3' EST AA783035 clone 1375023"  
repeat\_region complement(3220. 3367)  
/note="PR5 repeat: matches 288. 142 of consensus"  
repeat\_region complement(3265. 3898)  
/note="PR5 repeat: matches 2427. 1800 of consensus"  
repeat\_region complement(3900. 3945)  
/note="PR5 repeat: matches 1397. 1353 of consensus"  
3987. 4286  
/note="match: 253809 STS containing (CA) repeat"  
4102. 4127  
/note="13 copies of AC 100% conserved; differs from  
253809"  
5110. 5398  
/note="AluB repeat: matches 12. 302 of consensus"  
complement(5436. 5576)  
/note="MR2 repeat: matches 146. 4 of consensus"  
5606. 5937  
/note="MR2 repeat: matches 8. 344 of consensus"  
complement(6655. 7102)  
/note="match: GSS B82054"  
complement(7192. 7423)  
/note="11PA10 repeat: matches 900. 662 of consensus"  
9003. 9267  
/note="MR repeat: matches 2. 262 of consensus"  
complement(13451. 13902)  
/note="MT1D repeat: matches 505. 44 of consensus"  
14327. 14444  
/note="MR repeat: matches 47. 164 of consensus"  
complement(14857. 14933)  
/note="MR5A repeat: matches 179. 37 of consensus"  
15170. 15209  
/note="10 copies 4 mer tcat 83% conserved"  
complement(19889. 20167)  
/note="MR33 repeat: matches 302. 26 of consensus"  
24403. 24428  
/note="13 copies 2 mer tt 92% conserved"  
24433. 24855  
/note="MT2\_internal repeat: matches 1739. 2174 of  
consensus"  
complement(24871. 25777)  
/note="L1PA15 repeat: matches 904. 1 of consensus"  
complement(25626. 26986)  
/note="L1 repeat: matches 5390. 3989 of consensus"  
26985. 27537  
/note="L1 repeat: matches 2116. 2707 of consensus"  
complement(27553. 27820)  
/note="Alu repeat: matches 301. 34 of consensus"  
27822. 29084  
/note="L1 repeat: matches 2700. 3980 of consensus"  
29060. 29679  
/note="MT2\_internal repeat: matches 2145. 2774 of  
consensus"  
29763. 30148  
/note="STD repeat: matches 3. 394 of consensus"  
30148. 30776  
/note="MT2\_internal repeat: matches 2859. 3495 of

```

repeat_region      consensus
30922..31355
/ote="MIR2 internal repeat: matches 3710. .4142 of
consensus"
repeat_region      31659..32910
/ote="L1 repeat: matches 4149. .5390 of consensus"
32755..33606
/ote="LIMB6 repeat: matches 1. .912 of consensus"
repeat_region      complement(34384. .34589)
/ote="MIR repeat: matches 247. .32 of consensus"
35440..35544
/ote="MIR repeat: matches 4. .109 of consensus"
36090..36149
/ote="MIR repeat: matches 146. .48 of consensus"
<37067..37216
/ote="MIR repeat: matches 297. .2 of consensus"
/ote="AluSx repeat: matches 3794. .37994)
complement(37694. .37994)
/ote="AluSg repeat: matches 300. .1 of consensus"
38661..40016
/ote="match: multiple ESTs; match: EST Z38206 clone
c-02h10; Paired with EST Z41912 matching this clone;
match: EST Z41912 clone c-02h10; Paired with EST Z38206
matching this clone"
38661..38857
/ote="match: SRS G05514"
<40911..>42821
/ote="match: multiple ESTs; match: N50346 R49495 H21114
AA550242 H46781; match: H29818 F04199 H18678 H50505
F07947; match: H05281 F02238 N51569 H46205 H40284; match:
H46206 AA968020 R53041 R16141 H21157; match: Z42769 R53128
H29905 H51336 AA332574; match: R35107 AA968020 F07947
H18769 AA350243"
40991..>41471
/ote="match: SRS G24255"
41587..41620
/ote="17 copies 2 mer tg 85% conserved"
prim_transcript    complement(44566. .47136)
/ote="match: 5' EST R15400 clone 53057"
repeat_region      complement(46334. .46442)
/ote="MIR repeat: matches 157. .48 of consensus"
repeat_region      complement(46542. .46643)
/ote="MIR repeat: matches 145. .41 of consensus"
47141..47466
/ote="match: 3' EST AA985139 clone IMAGE:1616277"
repeat_region      complement(50606. .50641)
/ote="MIR2 repeat: matches 143. .108 of consensus"
50807..50922
/ote="MIR2 repeat: matches 9. .128 of consensus"
50878..50920
/ote="MIR repeat: matches 206. .248 of consensus"
/ote="MIR repeat: matches 192. .52 of consensus"
/ote="MIR repeat: matches 150. .81 of consensus"
/ote="MIR repeat: matches 146. .18 of consensus"
complement(56554. .56855)
/ote="AluY repeat: matches 301. .1 of consensus"
/ote="AluY repeat: matches 300. .1 of consensus"
complement(57507. .57630)
/ote="MIR repeat: matches 187. .66 of consensus"
complement(57973. .59073)
/ote="L1 repeat: matches 5047. .3929 of consensus"
complement(60809. .61101)
/ote="AluSx repeat: matches 293. .1 of consensus"
complement(62404. .64830)
/ote="TIGR1 repeat: matches 2418. .1 of consensus"
64848..64979

```

```

repeat_region      /ote="MIR2 repeat: matches 15. .146 of consensus"
65656..65790
/ote="MIR2 repeat: matches 4. .144 of consensus"
misc_feature       <65829..66276
/ote="match: GSS A0002124"
repeat_region      67773..67839
/ote="MIR repeat: matches 79. .146 of consensus"
68924..68987
/ote="32 copies 2 mer gt 83% conserved"
repeat_region      69506..70033
/ote="L1 repeat: matches 4843. .5389 of consensus"
69893..70891
/ote="LIMB6 repeat: matches 1. .1047 of consensus"

Query Match      91.6%; Score 17.4; DB 9; Length 100418;
Best Local Similarity 94.7%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 CTGGGGGCGATTATTCAC 19
Db      90926 CTGGGCTCAGTATTCAC 90908

RESULT 3
AC083819/c      AC083819      102165 bp      DNA      linear      HTG 16-OCT-2001
LOCUS      Mus musculus chromosome 1 clone RP23-285F20, *** SEQUENCING IN
DEFINITION      PROGRESS ***; 7 unordered pieces.
ACCESSION      AC083819
VERSION      AC083819
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE      Mus musculus.
ORGANISM      Mus musculus.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
1 (bases 1 to 102165)
Metzger,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okwuonu,G., Carllock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonlin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Fortum-Tansey,J., Gill,R.,
Gorrell,J.H., Gunatane,P., Haller,G., Hernandez,J., Hognes,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovay,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mel,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,
Worley,K. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 102165)
Worley,K.C.
Direct Submission
Submitted (01-OCT-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 16, 2000 this sequence version replaced gi:10801956.
----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: RP23-285F20
Sequencing vector: M13; 108821
Chemistry: Dye-primer Bodipy: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 95179 bases at least Q40

```

Consensus quality: 98007 bases at least Q30  
Consensus quality: 99497 bases at least Q20  
Estimated insert size: 100397; sum-of-coverage  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 2.9x in Q20 bases; sum-of-coverage estimation

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently  
consists of 7 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 24502: contig of 24502 bp in length  
\* 24503 24602: gap of unknown length  
\* 24603 40809: contig of 16207 bp in length  
\* 40810 40909: gap of unknown length  
\* 40910 63389: contig of 22480 bp in length  
\* 63390 63489: gap of unknown length  
\* 63490 78501: contig of 15012 bp in length  
\* 78502 78601: gap of unknown length  
\* 78602 89787: contig of 11186 bp in length  
\* 89788 89887: gap of unknown length  
\* 89888 96681: contig of 6794 bp in length  
\* 96682 96781: gap of unknown length  
\* 96782 102165: contig of 5384 bp in length.

## FEATURES

Location/Qualifiers  
1..102165  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="1"  
/clone="RP23-285F20"

BASE COUNT 30175 a 21419 c 20896 g 29070 t 605 others

## ORIGIN

Query Match 91.6%; Score 17.4; DB 2; Length 102165;  
Best Local Similarity 94.7%; Pred. No. 69;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTGGGGGCGATTATGCAC 19  
||||| |||||||||  
DB 82849 CTGGGGGCGATTATGCAC 82831

RESULT 4  
AC069000 158149 bp DNA linear HTG 12-AUG-2000  
LOCUS Homo sapiens chromosome 1 clone RP11-114G14, WORKING DRAFT  
DEFINITION  
AC069000  
AC069000.2 GI:8468957  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 158149)  
Waterston, R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 158149)  
Waterston, R.H.  
Direct Submission  
Submitted (16-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Jun 12, 2000 this sequence version replaced gi:7839915.

REFERENCE  
AUTHORS  
JOURNAL  
TITLE  
REFERENCE  
AUTHORS  
JOURNAL  
COMMENT  
On Jun 12, 2000 this sequence version replaced gi:7839915.

Center: Washington University Genome Sequencing Center

Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H.NH0114G14  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Sequencing vector: Plasmid; 0%  
Chemistry: Dye-primer ET; 100% of reads  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 151997 bases at least Q40  
Consensus quality: 153927 bases at least Q30  
Consensus quality: 155094 bases at least Q20  
Insert size: 126000; agarose-fp  
Insert size: 156749; sum-of-coverage  
Quality coverage: 5.73 in Q20 bases; agarose-fp  
Quality coverage: 4.63 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently  
consists of 15 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 1930: contig of 1930 bp in length  
\* 1931 2030: gap of unknown length  
\* 2031 3803: contig of 1773 bp in length  
\* 3804 3903: gap of unknown length  
\* 3904 5566: contig of 1663 bp in length  
\* 5567 5666: gap of unknown length  
\* 5667 10320: contig of 4654 bp in length  
\* 10321 10420: gap of unknown length  
\* 10421 16200: contig of 5780 bp in length  
\* 16201 16300: gap of unknown length  
\* 16301 23791: contig of 7491 bp in length  
\* 23792 23891: gap of unknown length  
\* 23892 32610: contig of 8719 bp in length  
\* 32611 32710: gap of unknown length  
\* 32711 43891: contig of 11181 bp in length  
\* 43892 43991: gap of unknown length  
\* 43992 53368: contig of 9377 bp in length  
\* 53369 53468: gap of unknown length  
\* 53469 66543: contig of 13075 bp in length  
\* 66544 66643: gap of unknown length  
\* 66644 80033: contig of 13390 bp in length  
\* 80034 80133: gap of unknown length  
\* 80134 94139: contig of 14006 bp in length  
\* 94140 94239: gap of unknown length  
\* 94240 107329: contig of 13090 bp in length  
\* 107330 107429: gap of unknown length  
\* 107430 131992: contig of 24363 bp in length  
\* 131993 132092: gap of unknown length  
\* 132093 158149: contig of 26057 bp in length.

## FEATURES

Location/Qualifiers  
1..158149  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-114G14"

misc\_feature  
1..1930  
/note="assembly\_name:Contig2"  
misc\_feature  
2031..3803  
/note="assembly\_name:Contig3"  
misc\_feature  
3904..5566  
/note="assembly\_name:Contig4"  
misc\_feature  
5667..10320  
/note="assembly\_name:Contig5"  
misc\_feature  
10421..16200  
/note="assembly\_name:Contig6"  
misc\_feature  
16301..23791  
/note="assembly\_name:Contig7"

```
misc_feature 23892..32610 /note="assembly_name:Contig8"
misc_feature 32711..43891 /note="assembly_name:Contig9"
misc_feature 43992..53368 /note="assembly_name:Contig10"
misc_feature 53469..66543 /note="assembly_name:Contig11"
misc_feature 66644..80033 /note="assembly_name:Contig12"
misc_feature 80134..94139 /note="assembly_name:Contig13"
misc_feature 94240..107329 /note="assembly_name:Contig14"
misc_feature 107430..131992 /note="assembly_name:Contig15"
misc_feature 132093..158149 /note="assembly_name:Contig16"
BASE COUNT 47784 a 31862 c 32435 g 44661 t 1407 others
ORIGIN
```

```
Query Match 91.6%; Score 17.4; DB 2; Length 158149;
Best Local Similarity 94.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 CTGGGGGAGTATTGTCAC 19
    ||||| ||||| ||||| |||||
Db 71696 CTGGGGTCACTATTGTCAC 71714
```

```
RESULT 5
AL445989 LOCUS AL445989 163218 bp DNA linear PRI 01-MAR-2001
DEFINITION Human DNA sequence from clone RP11-473M10 on chromosome 13,
ACCESSION AL445989 complete sequence.
VERSION AL445989.17 GI:13184381
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 163218)
REFERENCE
AUTHORS Kay.M.
TITLE Direct Submision
JOURNAL Submitted (01-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Mar 2, 2001 this sequence version replaced gi:12964329.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; SW:
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
Chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
```

RP11-473M10 is from the library RCI1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACE3.6  
This sequence is the entire insert of clone RP11-473M10. The true right end of clone RP11-671L7 is at 29875 in this sequence.

## FEATURES

```
source
    Location/Qualifiers
        1..163218
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="13"
            /clone="RP11-473M10"
            /clone_1lb="RP11-11.2"
        1..564
            /note="MER41B repeat: matches 3..572 of consensus"
        565..667
            /note="Aluub repeat: matches 3..104 of consensus"
        716..959
            /note="MER21B repeat: matches 232..670 of consensus"
        2000..2311
            /note="AluSq repeat: matches 1..236 of consensus"
        2611..2678
            /note="L1PA13 repeat: matches 6076..6143 of consensus"
        2679..2934
            /note="AluX repeat: matches 29..285 of consensus"
        2935..4792
            /note="L1PA13 repeat: matches 4206..6076 of consensus"
        4766..6224
            /note="L1PA13 repeat: matches 2769..4238 of consensus"
        6810..7104
            /note="AluSq repeat: matches 4..298 of consensus"
        7108..7193
            /note="43 copies 2 mer at 628 conserved"
        7545..7990
            /note="L1M3a repeat: matches 213..653 of consensus"
        8294..8666
            /note="L1M3a repeat: matches 745..983 of consensus"
        9017..10160
            /note="L1M2 repeat: matches 900..2373 of consensus"
        10261..10546
            /note="AluX repeat: matches 1..286 of consensus"
        12312..12547
            /note="L2 repeat: matches 2274..2522 of consensus"
        13764..14560
            /note="L1M6a repeat: matches 5518..6290 of consensus"
        14579..14648
            /note="L2 repeat: matches 2679..2748 of consensus"
        16707..16916
            /note="L1R16C repeat: matches 173..378 of consensus"
        17582..17630
            /note="MER89 repeat: matches 469..522 of consensus"
        18171..18228
            /note="29 copies 2 mer at 708 conserved"
        21313..21346
            /note="17 copies 2 mer ac 824 conserved"
        21420..21545
            /note="63 copies 2 mer at 764 conserved"
        21549..21817
            /note="AluX repeat: matches 21..294 of consensus"
        22448..22484
            /note="MIR repeat: matches 216..251 of consensus"
        23335..23400
            /note="2 copies 33 mer 954 conserved"
        24675..25831
            /note="L2 repeat: matches 748..2215 of consensus"
        25832..26127
            /note="AluSc repeat: matches 1..301 of consensus"
        26128..26589
            /note="L2 repeat: matches 344..748 of consensus"
        27049..27360
            /note="Alu repeat: matches 1..305 of consensus"
        27920..28006
            /note="MER57B repeat: matches 300..386 of consensus"
```



```

repeat_region 28192..28338
/note="MER33 repeat: matches 10..151 of consensus"
repeat_region 30716..31073
/note="179 copies 2 mer tt 54% conserved"
repeat_region 32832..33392
/note="LTR2CA repeat: matches 1..508 of consensus"
repeat_region 33643..33753
/note="LTR16C repeat: matches 262..378 of consensus"
repeat_region 34208..34245
/note="19 copies 2 mer aa 78% conserved"
repeat_region 34492..34884
/note="HERV.L repeat: matches 3334..3634 of consensus"
repeat_region 42139..42202
/note="MIR repeat: matches 79..142 of consensus"
repeat_region 42232..42383
/note="76 copies 2 mer ta 63% conserved"
repeat_region 43902..43926
/note="L1 repeat: matches 4804..4829 of consensus"
repeat_region 43927..44474
/note="L1MA8 repeat: matches 5733..6287 of consensus"
repeat_region 44475..44631
/note="L1 repeat: matches 4658..4804 of consensus"
repeat_region 45347..45650
/note="AluJo repeat: matches 4..305 of consensus"
repeat_region 46847..46940
/note="47 copies 2 mer tt 62% conserved"
repeat_region 47049..47110
/note="HALL repeat: matches 706..761 of consensus"
repeat_region 48421..48470
/note="75 copies 2 mer ac 90% conserved"
repeat_region 49812..50212
/note="LTR1A1 repeat: matches 1..365 of consensus"
repeat_region 52787..52850
/note="32 copies 2 mer ta 76% conserved"
repeat_region 53040..53341
/note="AluJ repeat: matches 1..303 of consensus"
repeat_region 54732..55698
/note="L1PA8 repeat: matches 4155..6158 of consensus"
repeat_region 56966..57975
/note="L1PA8 repeat: matches 2725..3982 of consensus"
repeat_region 57976..58471
/note="L2 repeat: matches 2268..2748 of consensus"
repeat_region 60694..60783
/note="L2 repeat: matches 2317..2407 of consensus"
repeat_region 61514..61934
/note="MER31A repeat: matches 3..429 of consensus"
repeat_region 61944..62082
/note="L2 repeat: matches 1770..1915 of consensus"
repeat_region 62161..62453
/note="AluJb repeat: matches 5..293 of consensus"
repeat_region 62568..62963
/note="L1M repeat: matches 2131..2549 of consensus"
repeat_region 63150..63533
/note="L1Mc repeat: matches 1516..1906 of consensus"
repeat_region 63638..63689
/note="MER47 repeat: matches 143..194 of consensus"
repeat_region 63687..63744
/note="MER47 repeat: matches 2244..2203 of consensus"
repeat_region 63775..63981
/note="L1Mc repeat: matches 1278..1497 of consensus"
repeat_region 63986..64486
/note="L1MC2 repeat: matches 5838..6330 of consensus"
repeat_region 64487..64834
/note="LTR1A1 repeat: matches 1..365 of consensus"
repeat_region 64835..65482
/note="L1MC2 repeat: matches 5164..5838 of consensus"
repeat_region 65489..65820
/note="L1MB8 repeat: matches 5810..6141 of consensus"
repeat_region 65883..66146
/note="L1MD repeat: matches 973..1251 of consensus"
repeat_region 66884..67288
/note="WSTB repeat: matches 1..426 of consensus"
repeat_region 67573..67853

```

```

/note="AluJSP repeat: matches 1..287 of consensus"
repeat_region 69319..69598
/note="LTR1B repeat: matches 54..364 of consensus"
repeat_region 71293..71849
/note="L1MA9 repeat: matches 4999..5555 of consensus"
repeat_region 71850..72551
/note="L1MA9 repeat: matches 5608..6308 of consensus"
repeat_region 75195..75277
/note="L2 repeat: matches 2660..2750 of consensus"
repeat_region 76986..77091
/note="L1PA3 repeat: matches 6041..6146 of consensus"
repeat_region 77667..77864
/note="99 copies 2 mer aa 56% conserved"
repeat_region 78135..80473
/note="TIGER1 repeat: matches 1..2407 of consensus"

Query Match 91.6%; Score 17.4; DB 9; Length 163218;
Best Local Similarity 94.7%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTGGGCGCAGTATTGCAC 19
DB 8349 CTGGGCGCAGTATTGCC 8367

RESULT 6
LOCUS AC016061
DEFINITION Homo sapiens clone RP11-6N9, WORKING DRAFT SEQUENCE, 15 unordered
pieces:
ACCESSION AC016061
VERSION AC016061.3 GI:7239536
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 171012)
Homo sapiens chromosome, clone RP11-6N9
Unpublished
2 (bases 1 to 171012)
Bairren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckert,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeBella,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Gallagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Liu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testafaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 14, 2000 this sequence version replaced g1:6705863.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: I2912
Center clone name: 6_N_9
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads

```

Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 161177 bases at least Q40  
 Consensus quality: 166152 bases at least Q30  
 Consensus quality: 168156 bases at least Q20  
 Insert size: 151000; agarose-ff  
 Insert size: 169612; sum-of-ctrls  
 Quality coverage: 4.9 in Q20 bases; agarose-ff  
 Quality coverage: 4.3 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently  
 consists of 15 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 1250: contig of 1250 bp in length  
 1251 1350: gap of 100 bp  
 1351 2445: contig of 1095 bp in length  
 2446 2545: gap of 100 bp  
 2546 3546: contig of 1001 bp in length  
 3547 3646: gap of 100 bp  
 3647 5950: contig of 2304 bp in length  
 5951 6050: gap of 100 bp  
 6051 9159: contig of 3109 bp in length  
 9160 9259: gap of 100 bp  
 9260 11925: contig of 2666 bp in length  
 11926 12025: gap of 100 bp  
 12026 16254: contig of 4229 bp in length  
 16255 16354: gap of 100 bp  
 16355 21948: contig of 5594 bp in length  
 21949 22048: gap of 100 bp  
 22049 29743: contig of 7695 bp in length  
 29744 29843: gap of 100 bp  
 29844 38782: contig of 8939 bp in length  
 38783 38882: gap of 100 bp  
 38883 53976: contig of 15094 bp in length  
 53977 54076: gap of 100 bp  
 54077 71196: contig of 17120 bp in length  
 71197 71296: gap of 100 bp  
 71297 101039: contig of 29743 bp in length  
 101040 101139: gap of 100 bp  
 101140 129583: contig of 28444 bp in length  
 129584 129684: gap of 100 bp  
 129684 171012: contig of 41329 bp in length.

## FEATURES

source  
 1. 171012  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP11-6N9"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1. 1250  
 /note="assembly\_fragment"  
 misc\_feature  
 1351..2445  
 /note="assembly\_fragment"  
 misc\_feature  
 2546..3546  
 /note="assembly\_fragment"  
 misc\_feature  
 3647..5950  
 /note="assembly\_fragment"  
 misc\_feature  
 6051..9159  
 /note="assembly\_fragment"  
 misc\_feature  
 9260..11925  
 /note="assembly\_fragment"  
 misc\_feature  
 12026..16254  
 /note="assembly\_fragment"  
 misc\_feature  
 16355..21948  
 /note="assembly\_fragment"

misc\_feature 22049..29743  
 /note="assembly\_fragment"  
 misc\_feature 29844..38782  
 /note="assembly\_fragment"  
 misc\_feature 38883..53976  
 /note="assembly\_fragment"  
 misc\_feature 54077..71196  
 /note="assembly\_fragment"  
 misc\_feature 71297..101039  
 /note="assembly\_fragment"  
 misc\_feature 101140..129583  
 /note="assembly\_fragment"  
 misc\_feature 129684..171012  
 /note="assembly\_fragment"  
 BASE COUNT 55604 a 29837 c 29224 g 54946 t 1401 others  
 ORIGIN

Query Match 91.6% Score 17.4; DB 2; Length 171012;  
 Best Local Similarity 94.7% Pred. No. 67;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CTGGGGCAGTATTGCAC 19  
 |||||  
 Db 21022 CTGGGGCAGTATTGCC 21040

RESULT 7  
 AC129875/c 202509 bp DNA linear HTG 04-AUG-2002

DEFINITION Rattus norvegicus clone CH230-11363, \*\*\* SEQUENCING IN PROGRESS

LOCUS \*\*\* 65 unordered pieces.

AC129875 AC129875.1 GI:22095101

VERSION HTG: HTGS-PHASE1.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 202509)

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,  
 Alshrocks,S.L., Amartunge,H.C., Are,J.R., Ayala,M., Banks,T.,  
 Barbieri,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
 Bouck,U., Bowls,S., Brlewa,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhaq,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,E.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douhalwalte,K.J., Draper,H., Dugan-Rocha,S., Dublin,K.J.,  
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frazz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
 Homsi,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kratochvic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,  
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhney,E., Mcleod,M.P., Meador,M., Mel,G., Metzger,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,S., Ogun,M., Okunonu,G.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunonu,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Plickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rives,M., Rojas,A., Rojokokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,  
 Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Taney, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, O.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.  
Direct Submission  
Unpublished  
2 (bases 1 to 202509)

## COMMENT

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: CGOB  
Center clone name: CH230-11363  
----- Summary Statistics  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap: version 0.990329  
Consensus quality: 142504 bases at least Q40  
Consensus quality: 151466 bases at least Q30  
Consensus quality: 158518 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 65 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1775: contig of 1775 bp in length  
\* 1776 1875: gap of unknown length  
\* 1876 3503: contig of 1628 bp in length  
\* 3504 3603: gap of unknown length  
\* 3604 4952: contig of 1349 bp in length  
\* 4953 5052: gap of unknown length  
\* 5053 6252: contig of 1200 bp in length  
\* 6253 6352: gap of unknown length  
\* 6353 7765: contig of 1413 bp in length  
\* 7766 7865: gap of unknown length  
\* 7866 9274: contig of 1409 bp in length  
\* 9275 9374: gap of unknown length  
\* 9375 10620: contig of 1246 bp in length  
\* 10621 10720: gap of unknown length  
\* 10721 11906: contig of 1186 bp in length  
\* 11907 12006: gap of unknown length  
\* 12007 13579: contig of 1573 bp in length  
\* 13580 13679: gap of unknown length  
\* 13680 15828: contig of 2149 bp in length  
\* 15829 15928: gap of unknown length  
\* 15929 17956: contig of 2028 bp in length  
\* 17957 18056: gap of unknown length  
\* 18057 19850: contig of 1794 bp in length  
\* 19851 19950: gap of unknown length  
\* 19951 22134: contig of 2184 bp in length  
\* 22135 22234: gap of unknown length  
\* 22235 23686: contig of 1452 bp in length  
\* 23687 23786: gap of unknown length  
\* 23787 25813: contig of 2027 bp in length  
\* 25814 25913: gap of unknown length  
\* 25914 27123: contig of 1210 bp in length  
\* 27123 27223: gap of unknown length  
\* 27223 27223: gap of unknown length

27224 29041: contig of 1818 bp in length  
\* 29042 29141: gap of unknown length  
\* 29142 30795: contig of 1654 bp in length  
\* 30796 30895: gap of unknown length  
\* 30896 33092: contig of 2197 bp in length  
\* 33093 33192: gap of unknown length  
\* 33193 35125: contig of 1933 bp in length  
\* 35126 35225: gap of unknown length  
\* 35226 37231: contig of 2006 bp in length  
\* 37232 37331: gap of unknown length  
\* 37332 39368: contig of 2037 bp in length  
\* 39369 41641: gap of unknown length  
\* 41642 41741: gap of unknown length  
\* 41742 43905: contig of 2164 bp in length  
\* 43906 44005: gap of unknown length  
\* 44006 45362: contig of 1357 bp in length  
\* 45363 45462: gap of unknown length  
\* 45463 47720: contig of 2258 bp in length  
\* 47721 47820: gap of unknown length  
\* 47821 49233: contig of 1413 bp in length  
\* 49234 49333: gap of unknown length  
\* 49334 51083: contig of 1750 bp in length  
\* 51084 51183: gap of unknown length  
\* 51184 53500: contig of 2317 bp in length  
\* 53501 53600: gap of unknown length  
\* 53601 55235: contig of 1635 bp in length  
\* 55236 55335: gap of unknown length  
\* 55336 57601: contig of 2266 bp in length  
\* 57602 57701: gap of unknown length  
\* 57702 60310: contig of 2609 bp in length  
\* 60311 60410: gap of unknown length  
\* 60411 63132: contig of 2722 bp in length  
\* 63133 63232: gap of unknown length  
\* 63233 65772: contig of 2540 bp in length  
\* 65773 65872: gap of unknown length  
\* 65873 68571: contig of 2699 bp in length  
\* 68572 68671: gap of unknown length  
\* 68672 70187: contig of 1516 bp in length  
\* 70188 70287: gap of unknown length  
\* 70288 72875: contig of 2688 bp in length  
\* 72876 73075: gap of unknown length  
\* 73076 74970: contig of 1895 bp in length  
\* 74971 75070: gap of unknown length  
\* 75071 78468: contig of 3398 bp in length  
\* 78469 78568: gap of unknown length  
\* 78569 80353: contig of 1785 bp in length  
\* 80354 80453: gap of unknown length  
\* 80454 83088: contig of 2635 bp in length  
\* 83089 83188: gap of unknown length  
\* 83189 87167: contig of 3979 bp in length  
\* 87168 87267: gap of unknown length  
\* 87268 89490: contig of 2223 bp in length  
\* 89491 89590: gap of unknown length  
\* 89591 92487: contig of 2897 bp in length  
\* 92488 92587: gap of unknown length  
\* 92588 95643: contig of 3056 bp in length  
\* 95644 95743: gap of unknown length  
\* 95744 98940: contig of 3187 bp in length  
\* 98941 99040: gap of unknown length  
\* 99041 101199: contig of 2159 bp in length  
\* 101200 101299: gap of unknown length  
\* 101300 105196: contig of 3897 bp in length  
\* 105197 105296: gap of unknown length  
\* 105297 108314: contig of 3018 bp in length  
\* 108315 108414: gap of unknown length  
\* 108415 110704: contig of 2280 bp in length  
\* 110705 110804: gap of unknown length  
\* 110805 113320: contig of 2516 bp in length  
\* 113321 113420: gap of unknown length  
\* 113421 117654: contig of 4234 bp in length  
\* 117655 117754: gap of unknown length  
\* 117755 122449: contig of 4695 bp in length

```

* 122450 122549: gap of unknown length
* 122550 127482: contig of 4933 bp in length
* 127583 127582: gap of unknown length
* 127583 131394: contig of 3812 bp in length
* 131395 131494: gap of unknown length
* 131495 135752: contig of 4258 bp in length
* 135753 135852: gap of unknown length
* 135853 141135: contig of 5283 bp in length
* 141136 141235: gap of unknown length

Query Match      91.6%  Score 17.4;  DB 2;  Length 202509;
Best Local Similarity 94.7%  Pred. No. 67;
Matches 18;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Oy 1 CTGGGGGACGTTATTGCAC 19
    ||||| ||||| ||||| |||||
Db 151028 CTGGGGTCAGTTATTGCAC 151010

RESULT 8
AC098629/c
LOCUS
DEFINITION
AC098629 168435 bp DNA linear HTG 23-JUL-2002
Rattus norvegicus clone CH230-108122, *** SEQUENCING IN PROGRESS
AC098629
AC098629.4 GI:21239840
HTG: HTGS_PHASE1.
SOURCE
ORGANISM
Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 168435)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alspbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Burke,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Hornai,F., Howard,S., Huber,J., Hulik,S., Hume,D., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,D., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Louisgeed,H.,
Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maneshwarl,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nockenko,S., Ogih,M., Okunou,G.,
Oregunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., RojudoKan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Shaver,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,K., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wlaczkyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (27-Oct-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 168435)
Worley,K.C.
Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 29, 2002 this sequence version replaced gi:17973473.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHXD
Center clone name: CH230-108122
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 105964 bases at least Q40
Consensus quality: 111011 bases at least Q30
Consensus quality: 115429 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1058: contig of 1058 bp in length
* 1059 1158: gap of unknown length
* 1159 2246: contig of 1088 bp in length
* 2247 2346: gap of unknown length
* 2347 3914: contig of 1568 bp in length
* 3915 4015: gap of unknown length
* 4015 5410: contig of 1396 bp in length
* 5411 5511: gap of unknown length
* 5511 7015: contig of 1505 bp in length
* 7016 7116: gap of unknown length
* 7116 8297: contig of 1182 bp in length
* 8298 8397: gap of unknown length
* 8398 9457: contig of 1060 bp in length
* 9458 9557: gap of unknown length
* 9558 11326: contig of 1769 bp in length
* 11327 11426: gap of unknown length
* 11427 12468: contig of 1042 bp in length
* 12469 12568: gap of unknown length
* 12569 13856: contig of 1288 bp in length
* 13857 13956: gap of unknown length
* 13957 15471: contig of 1515 bp in length
* 15472 15571: gap of unknown length
* 15572 16585: contig of 1014 bp in length
* 16586 16685: gap of unknown length
* 16686 17938: contig of 1253 bp in length
* 17939 18038: gap of unknown length
* 18039 19446: contig of 1408 bp in length
* 19447 19546: gap of unknown length
* 19547 20676: contig of 1130 bp in length
* 20677 22119: gap of unknown length
* 22120 22219: gap of unknown length

```

```

22220 23410: contig of 1191 bp in length
* 23411 23510: gap of unknown length
* 23511 24594: contig of 1084 bp in length
* 24595 24694: gap of unknown length
* 24695 25900: contig of 1206 bp in length
* 25901 26000: gap of unknown length
* 26001 27046: contig of 1046 bp in length
* 27047 27146: gap of unknown length
* 27147 28430: contig of 1284 bp in length
* 28431 28530: gap of unknown length
* 28531 29780: contig of 1250 bp in length
* 29781 29880: gap of unknown length
* 29881 31611: contig of 1731 bp in length
* 31612 31712: gap of unknown length
* 31713 33655: contig of 1944 bp in length
* 33656 33755: gap of unknown length
* 33756 35174: contig of 1419 bp in length
* 35175 35274: gap of unknown length
* 35275 37331: contig of 2057 bp in length
* 37332 37431: gap of unknown length
* 37432 39444: contig of 2013 bp in length
* 39445 39544: gap of unknown length
* 39545 40865: contig of 1321 bp in length
* 40866 40965: gap of unknown length
* 40966 43515: contig of 2550 bp in length
* 43516 43615: gap of unknown length
* 43616 45690: contig of 2075 bp in length
* 45691 45790: gap of unknown length
* 45791 46803: contig of 1013 bp in length
* 46804 46903: gap of unknown length
* 46904 49441: contig of 2538 bp in length
* 49442 49541: gap of unknown length
* 49542 51912: contig of 2371 bp in length
* 51913 52012: gap of unknown length
* 52013 54865: contig of 2884 bp in length
* 54867 54966: gap of unknown length
* 54967 56315: contig of 1319 bp in length
* 56316 56415: gap of unknown length
* 56416 59301: contig of 2886 bp in length
* 59302 59401: gap of unknown length
* 59402 62164: contig of 2763 bp in length
* 62165 62264: gap of unknown length
* 62265 64647: contig of 2383 bp in length
* 64648 64747: gap of unknown length
* 64748 66818: contig of 2071 bp in length
* 66819 66918: gap of unknown length
* 66919 69765: contig of 2847 bp in length
* 69766 69865: gap of unknown length
* 69866 73674: contig of 3809 bp in length
* 73675 73774: gap of unknown length
* 73775 76147: contig of 2373 bp in length
* 76148 76247: gap of unknown length
* 76248 79383: contig of 3136 bp in length
* 79384 79483: gap of unknown length
* 79484 82525: contig of 3042 bp in length
* 82526 82625: gap of unknown length
* 82626 86477: contig of 3852 bp in length
* 86478 86577: gap of unknown length
* 86578 88928: contig of 2351 bp in length
* 88929 89028: gap of unknown length
* 89029 92657: contig of 3629 bp in length
* 92658 92757: gap of unknown length
* 92758 95689: contig of 2932 bp in length
* 95690 95789: gap of unknown length
* 95790 100042: contig of 4253 bp in length
* 100043 100142: gap of unknown length
* 100143 103701: contig of 3559 bp in length
* 103702 103801: gap of unknown length
* 103802 107288: contig of 3487 bp in length
* 107289 107388: gap of unknown length
* 107389 112161: contig of 4773 bp in length
* 112162 112261: gap of unknown length
* 112262 115016: contig of 2755 bp in length

```

```

* 115017 115116: gap of unknown length
* 115117 120847: contig of 5731 bp in length
Query Match 89.5%; Score 17; DB 2; Length 168435;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TGCGGCGAGTTATTCGA 18
Db 156771 TGCGGCGAGTTATTCGA 156755

```

```

RESULT 9
AC117955/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-297H21, *** SEQUENCING IN PROGRESS
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 170238)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amaralunge,R.C., Are,J.R., Ayale,M., Banks,T.,
Barbala,J., Benton,J., Bimaye,K., Blackenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carion,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dabohne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Doutwate,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J.J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gottlieb,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Meheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Messey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabac,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Plums,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolte,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,H.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,O.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Unpublished
Direct Submission
2 (bases 1 to 170238)
Submitted (12-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

```

REFERENCE Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 170238)  
 AUTHORS Morley, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 14, 2002 this sequence version replaced gi:20258124.

Center: Baylor College of Medicine

Genome Center

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Project Information

Center project name: GTUN

Center clone name: CH230-297H21

Summary Statistics

Sequencing vector: plasmid:

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 129980 bases at least Q40

Consensus quality: 135538 bases at least Q30

Consensus quality: 139850 bases at least Q20

NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 43 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence.  
 as soon as it is available and the accession number will  
 be preserved.

1 1294: contig of 1294 bp in length  
 \* 1295 1394: gap of unknown length  
 \* 1395 2602: contig of 1208 bp in length  
 \* 2603 2702: gap of unknown length  
 \* 2703 4328: contig of 1626 bp in length  
 \* 4329 4428: gap of unknown length  
 \* 4429 6030: contig of 1602 bp in length  
 \* 6031 6130: gap of unknown length  
 \* 6131 7521: contig of 1391 bp in length  
 \* 7522 7621: gap of unknown length  
 \* 7622 8853: contig of 1232 bp in length  
 \* 8854 8953: gap of unknown length  
 \* 8954 10205: contig of 1252 bp in length  
 \* 10206 10305: gap of unknown length  
 \* 10306 11476: contig of 1171 bp in length  
 \* 11477 11576: gap of unknown length  
 \* 11577 13046: contig of 1470 bp in length  
 \* 13047 13146: gap of unknown length  
 \* 13147 14641: contig of 1495 bp in length  
 \* 14642 14741: gap of unknown length  
 \* 14742 17016: contig of 2275 bp in length  
 \* 17017 17116: gap of unknown length  
 \* 17117 19133: contig of 2017 bp in length  
 \* 19134 19233: gap of unknown length  
 \* 19234 20261: contig of 1028 bp in length  
 \* 20262 20361: gap of unknown length  
 \* 20362 22138: contig of 1777 bp in length  
 \* 22139 22238: gap of unknown length  
 \* 22239 24389: contig of 2151 bp in length  
 \* 24390 24489: gap of unknown length  
 \* 24490 26393: contig of 1904 bp in length  
 \* 26394 26493: gap of unknown length  
 \* 26494 28677: contig of 2184 bp in length  
 \* 28678 28777: gap of unknown length  
 \* 28779 30246: contig of 1469 bp in length  
 \* 30247 30346: gap of unknown length  
 \* 30347 32108: contig of 1662 bp in length  
 \* 32109 33927: contig of 1819 bp in length

33928 34027: gap of unknown length  
 \* 34028 36660: contig of 2633 bp in length  
 \* 36661 36760: gap of unknown length  
 \* 36761 38570: contig of 1810 bp in length  
 \* 38571 38670: gap of unknown length  
 \* 38671 41138: contig of 2468 bp in length  
 \* 41139 41238: gap of unknown length  
 \* 41239 43266: gap of unknown length  
 \* 43267 47194: contig of 3928 bp in length  
 \* 47195 47294: gap of unknown length  
 \* 47295 51037: contig of 3743 bp in length  
 \* 51038 51137: gap of unknown length  
 \* 51138 54201: contig of 2964 bp in length  
 \* 54202 56590: contig of 2389 bp in length  
 \* 56591 56690: gap of unknown length  
 \* 56691 59953: contig of 3263 bp in length  
 \* 59954 60053: gap of unknown length  
 \* 60054 64338: contig of 4285 bp in length  
 \* 64339 68405: gap of unknown length  
 \* 68406 68505: gap of unknown length  
 \* 68506 73009: contig of 4504 bp in length  
 \* 73010 73109: gap of unknown length  
 \* 73110 79117: contig of 6008 bp in length  
 \* 79118 79217: gap of unknown length  
 \* 79219 82852: contig of 3635 bp in length  
 \* 82853 82952: gap of unknown length  
 \* 82953 87619: contig of 4667 bp in length  
 \* 87620 87720: gap of unknown length  
 \* 87721 96467: contig of 8748 bp in length  
 \* 96468 96567: gap of unknown length  
 \* 96568 103197: contig of 6630 bp in length  
 \* 103198 103297: gap of unknown length  
 \* 103298 109986: contig of 6689 bp in length  
 \* 109987 120973: gap of unknown length  
 \* 120974 121073: contig of 10887 bp in length  
 \* 121074 130608: contig of 9535 bp in length  
 \* 130609 130708: gap of unknown length  
 \* 130709 143095: contig of 12387 bp in length  
 \* 143096 143195: gap of unknown length  
 \* 143196 153124: contig of 9929 bp in length  
 \* 153125 153224: gap of unknown length  
 \* 153225 170238: contig of 17014 bp in length.  
 Location/Qualifiers  
 1..170238  
 source  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /clone="CH230-297H21"

BASE COUNT 44877 a 35872 c 35293 g 45438 t 8758 others

ORIGIN

Query Match 89.5%; Score 17; DB 2; length 170238; 1

Best local Similarity 100.0%; Pred. No. 1.1e+02; 1

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TGGGGCAGTATTGCA 18

Db 105814 TGGGGCAGTATTGCA 105798

RESULT 10

AK091176 2785 bp mRNA linear PRI 15-JUL-2002

LOCUS Homo sapiens CDNA FLJ33857 f1s, clone CTONG2006004.

DEFINITION AK091176

ACCESSION AK091176

VERSION AK091176.1 GI:21749482

KEYWORDS oligo capping; f1s (full insert sequence).

SOURCE Homo sapiens tongue, tumor tissue CDNA to mRNA, clone\_11b:CTONG2

ORGANISM Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	AUTHORS	TITLE	JOURNAL	COMMENT
1	Nishi, T., Ota, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, Y., Isono, Y., Kawai, H., Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Takahashi, H., Kanda, K., Magatsuma, M., Murekawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Sugano, S., Nagahari, K., Masuno, Y., Nagai, K., and Isogai, T.	NEBO human cDNA sequencing project	Unpublished				
2	(bases 1 to 2785)	Isogai, T. and Yamamoto, J.	Direct Submission				
Submitted (04-JUL-2002)	Takao Isogai, FUJ Project(HRI Team); 2-6-7						
	Karusu-Kamatari, Kisarazu, Chiba 259-0812, Japan						
	(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)						
	NEBO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.						
FEATURES	source	1. 2785					
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	/clone="CTONG2006004"						
	/tissue_type="tongue, tumor tissue"						
	/clone_id="CTONG2"						
	/note="cloning vector: PME18SFL3"						
	204.. 2630						
	/note="unnamed protein product"						
	/codon_start=1						
	/protein_id="BAC03600.1"						
	/db_xref="GI:21749483"						
	/translation="MKYANILLSTYCRNEDAOYFIREVAEKLSTLADGETTEELVLEEGSTANISIOOEIKRLSESSSSOMALVOCRLHNMKLSLYLBCKAKAMDILPFIHLSOHNYHAPAEKVLIOYFSPVIDMILAPENIPVTSKMSDNOKCRPOEIOGSGWPTDPEILLCSSEPDSSMHMLVAVKQAPILSYLACLOQAKSISCLCVATITSEVDNATBEAGHITODSTEDTHTWEDLSVIRLTLLTKOKSTLIRGQLFFKDSPLLVSENEATCEMFENYKKEAEALTLFEKSLLETLNATVHPVTPAMWLEDVCFEPLKMLQCKQYELQELQTLVEREHLEFSDGPVKKLCILCOIKDLSIAINHTLITSYSIENLQHEGCSILERLQTDGOFALARVAELALPDVNLVYKETQOMOTIKHEDQSLKORADIVPMKCHENFKKNSISKSAFSPSTQAVHCEPTGQSSMEERHLLITTLGHMTAODVPMKLELELEIOIMCRITQHOHLGADNDEPFRFSNOITSSELSFDSASSEFSKLAALNTSKYLELNSLDSKETCEKRLDMKDESNLFLGRLLDGCHEARVCRVTRFYRNDVALVLCRALASGEASMEDELHEIALQASLEAEADIPRRARHSTSDHFKFNVTVSSNEVNTVNLVITSKLIHKNRCROYCLCYDLALLEGCSYTDVAADGEAMLRKILIASOQPDRCRAQAFISTGCLPDPVAVELVAEEVTRELLTSQGTGALPDPGSOPL"						
BASE COUNT	800 a	624 c	643 g	718 t			
ORIGIN							
Query Match	86.3%	Score 16.4	DB 9	Length 2785			
Best Local Similarity	94.4%	Pred. No. 3e+02					
Matches 17	Conservative 0	Mismatches 1	Indels 0	Gaps 0			
OY	1	CTGGGGGACGATTATGCA	18				
Db	677	CTGGGGGACACTTATTGCA	660				
RESULT 11	AB058743	7752 bp	mRNA	linear	PRI 05-JUL-2002		
LOCUS	AB058743						
DEFINITION	Homo sapiens mRNA for KIM1840 protein, partial cds.						
ACCESSION	AB058743						
KEYWORDS	AB058743.2	GI:20521987					

SOURCE	ORGANISM	REFERENCE AUTHORS TITLE	JOURNAL	MEDLINE PUBMED	REFERENCE AUTHORS TITLE	COMMENT FEATURES
Homio sapiens brain cDNA to mRNA, clone_1lib:pb1uescript1it SK plus	Homio sapiens	Nagase,T., Nakayama,M., Nakajima,D., Kikuno,R. and Ohara,O.		21245130		
clone:pf01011.	Homio sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		11347906		
		1		2 (bases 1 to 7752)		
		2 (bases 1 to 7752)				
		Chara,O., Nagase,T. and Kikuno,R.				
		Direct Submission				
		Submitted (27-Mar-2001) Osamu Ohara, Kazusa DNA Research Institute,				
		Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba				
		292-0812, Japan (E-mail:cdnaahfokazusa.or.jp,				
		URL:http://www.kazusa.or.jp/nuge, Tel:81-438-52-3913,				
		Fax:81-438-52-3914)				
		On May 9, 2002 this sequence version replaced gi:14017896.				
		Location/Qualifiers				
		1..7752				
		/organism="Homio sapiens"				
		/db_xref="taxon:9606"				
		/clone="pf01011"				
		/tissue-type="brain"				
		/clone_lib="pb1uescript1it SK plus"				
		/note="This sequence was replaced that of f19761 cDNA as				
		a representative cDNA sequence for				
		KIA1840.-vector:pb1uescript1it SK plus"				
		1..7752				
		/gene="KIA1840"				
		<1..7733				
		/gene="KIA1840"				
		/note="start codon is not identified"				
		/codon_start=2				
		/product="KIA1840 protein"				
		/protein_id="BAB47469.2"				
		/db_xref="gi:20521988"				
		/translation="MAAPEGYASAAGSGSMGTAMGRVLPMLTPVPAEAMGOLGS				
		/AQLTPQPAJLQSLTAAGSLQVLTLPSPGSGGRCLESPPHFLMEBSRNSPTTERK				
		KLALGENVELLIYFNMDKRCADATILSCSRALQKLDQDITSLSLRLTFLSS				
		NNTSLLEINKCVIIHIIFFERPERDAIVLNCETLIPQAVMDIIDTQCRGILVSS				
		LGMYIFPDVVGATVAHADVAHLHREKEDMNEQQPRAISFTSLASODLVAIVSS				
		SNSVAIVNLYEFROGHGHLCEERLEDELPGQKGVDELDPVNSAVNMKLAKESFO				
		DRSKAOLSLINETIKNSKLESCCAPMFOIILLESQSNHSTVSWAFIPDQIM				
		HGQVNYLQKHAKTSDPGRSKMHHSIDQERITLKCYSYGFPALETWYBERGYIT				
		TLWDELQGMCCSLGTRKCLPVDSSGQDCLFVLENGLSLLEGLQEBELNKLMLH				
		GSATVDTLCHLNGWGRCSIPHALLENGIENRQIDTVNFKLSKENTLFPSSKSSVS				
		QFDHLSHLYLRNVEELIPALDILCSAIRESESPQSHFSEQLNLTSLNNQKIK				
		LEITREDELHLOKVNIILSYINELRTFMKIPMKLTDAIDEDVHENVKVESNS				
		WKTSFPEVIALASALINNKIPAPQAFRRIDSHSAQKLEELIGLIVLNDLKNKIK				
		ASLEKLMGDKVQGLKICGYTNKNIRDLVILKEKDYENESREKRTIDVHVALE				
		LYLGRDNNQIOGSPRYKTEQDPFRKHSYVDSFLTEKQDERNNKDRHIVYALML				
		WDLTQESILPRLISPEYKSYSEPALMRVITAHMDNITLLMIGERFQTSVASSQ				
		NKMWLTIVDVINQMTSCNNYWRNELLQKLANGVFLASLEDEFCFLRLSRIGGV				
		DTLQVNYKTEKGWDFHSQFLTYCLEHSLQHLVLYVLDYCSLSENCPLFEKKEH				
		HPMEFIVQCVQVANSNLNLTDLPLKILFOASLANQOILIPRQASVSMLEGHIL				
		AT				
		YHLTOSLSPDPSRLFGMOSANLTALIDAGSHLPHSPDLYNKKVAYIERENFV				
		NGRSPAFGFTLVQELIKSKTKPKQIQOVQNEAYVIGLSHSPHYSIGACACYFEL				
		LDLSLAKNDKVNANIILSYKCRNEDAOISFIRSEVAELSLADGERTTDELVLLE				
		EGTANSIQOQEIIRLSESSQMALVVOFCPLNMKLSIYLRCACANMDLOFTIHS				
		QLANHRAPEVYSILOEYSPVIODLRLAFENMLPEVPSKMDQVNCNCPQELQSK				
		EMDTPEILLOCSSEPDHMLLVEAVQALQVILSYLASQGASALCLQVMIITS				
		EDNVATEAMGHIQDSTDHMTNLELSIWMTEILTRQSKTLIGLGFOLFEPDPS				
		MYEELCMFPNRYEAKRKLLEPKSLETTLNTATYKHPVPAWMLDEQVQELIKM				
		QOQCTQYBELGLDQLEVERHELSDEGPVKVLCILQILQKLSAINHTITTSSTK				
		LOCHCRSILERLQTDQGFALARAVALAEIVDVNLVLEITQEQVTKHIEQWSIKQ				
		RIDWPKCHENFKFKNSSIKSAASFTGQVAAQVDEHPGMSMERHLLIAGWML				
		QEDVWPLDKLEELKQIMICRTQHTTIGRNOEETPEPFRSRIISNGSEFSPSL				



```

FSLAALNTSKYLELNSLPSKPTCENRLDWMKEQESLNFLLIGRLDGDGVHSAVRVRY
FHYNDVALVJHCRALASGEASMEDEIHEIHALQSAELIEEPIDILRVSHTSS
LDSOKFVYVSSNENVTNLEVTLSKLNKNCROVCLCYDLAKEGCCTYVAADG
EMLRKIIASOOPDRCKRAQARTSTOGLKAEVAVAEVRELLTSOGSGHOMF
NPTESOTFLOLJTLICODRTGLGKMLDKISSVPHRESCTEBELLIAHCCTICOM
EELIRYLAQAAHMLTDNHLAPSEBYGIVYVLLLGICRYNEMTYIFDLIKKHFEVIMR
KLDPGSLKTLALDITKRCRPGDSKHNHIALCFSEWREIGENHEAARIDULKLES
QPMEDSLKGGHOLKOLLKALTLMDAESYAKDSVROAQHCOBLTTLITQIHFLN
TSCONTLMNGHKLMDICILALPREFYQASIVAEVFPDMAELIYQOYLKGFENYL
EEFKOORLKSIFEEISKRYKHOPTDWMENIKKLTYCEDVLYLYKLAVEHKEFYE
IYNVLKDPOGCGCKDLMLAG"
BASE COUNT      2256 a 1672 c 1721 g 2103 t
ORIGIN
Query Match      86.3%; Score 16.4; DB 9; Length 7752;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      1 CTGGGGCGAGTTATTCGA 18
        11111111111111111111
        4288 CTGGGGCGACTTATTCGA 4271
Db
RESULT 12
AF000195/c      AF000195      30102 bp      DNA      linear      INV 19-APR-2002
LOCUS      Caenorhabditis elegans cosmid C32F10, complete sequence.
DEFINITION
ACCESSION      AF000195
VERSION
KEYWORDS      HTG.
SOURCE
ORGANISM      Caenorhabditis elegans.
                Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
                Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
AUTHORS        Waterston, R.
TITLE          1 (bases 1 to 30102)
JOURNAL        Genome sequence of the nematode C. elegans: a platform for
                investigating biology. The C. elegans Sequencing Consortium
                Science 282 (5396), 2012-2018 (1998)
MEDLINE        99069613
PUBMED
REFERENCE
AUTHORS        Sammons, L. and Wohldmann, P.
TITLE          2 (bases 1 to 30102)
JOURNAL        The sequence of C. elegans cosmid C32F10
REFERENCE
AUTHORS        Waterston, R.
TITLE          3 (bases 1 to 30102)
JOURNAL        Unpublished (2001)
REFERENCE
AUTHORS        Waterston, R.
TITLE          Direct Submission
JOURNAL        Submitted (17-APR-1997) Genome Sequencing Center, Washington
                University
REFERENCE
AUTHORS        4 (bases 1 to 30102)
JOURNAL        Waterston, R.
TITLE          Direct Submission
JOURNAL        Submitted (08-MAY-1997) Genome Sequencing Center, Washington
                University
REFERENCE
AUTHORS        5 (bases 1 to 30102)
JOURNAL        Waterston, R.
TITLE          Direct Submission
JOURNAL        Submitted (31-DEC-1997) Department of Genetics, Washington
                University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
AUTHORS        6 (bases 1 to 30102)
JOURNAL        Waterston, R.
TITLE          Direct Submission
JOURNAL        Submitted (25-JUN-1998) Department of Genetics, Washington
                University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
AUTHORS        7 (bases 1 to 30102)
JOURNAL        Waterston, R.
TITLE          Direct Submission
JOURNAL        Submitted (05-JAN-2000) Department of Genetics, Washington
                University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
AUTHORS        8 (bases 1 to 30102)
JOURNAL        Waterston, R.
TITLE          Direct Submission

```

```

JOURNAL        Submitted (14-MAR-2001) Department of Genetics, Washington
                University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
AUTHORS        9 (bases 1 to 30102)
JOURNAL        Waterston, R.
TITLE          Direct Submission
JOURNAL        Submitted (05-OCT-2001) Department of Genetics, Washington
                University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
                Louis, MO 63110, USA
REFERENCE
AUTHORS        10 (bases 1 to 30102)
JOURNAL        Waterston, R.
TITLE          Direct Submission
JOURNAL        Submitted (19-APR-2002) Department of Genetics, Washington
                University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
                Louis, MO 63110, USA
COMMENT        On Dec 31, 1997 this sequence version replaced gi:1946996.
                Submitted by:
                Genome Sequencing Center
                Department of Genetics, Washington University
                St. Louis, MO 63110, USA, and
                Sanger Centre, Hinxton Hall
                Cambridge CB10 1HQ, England
                email: rwenematode.wustl.edu and jesse@sanger.ac.uk

```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., paired quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this cosmid sequence and its analysis see: {www.wormbase.org/db/seq/sequence?name=C32F10;class=Sequence}

#### NEIGHBORING COSMID INFORMATION

The 5' cosmid is B0414, 5900 bp overlap; the 3' cosmid is F33D1, 1100 bp overlap. Actual start of this cosmid is at base position 1 of C32F10.

#### NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFINDER (P. Green and L. Hillier, personal communication), the large scale EST projects of Yujl Kohara (http://www.ddbj.nig.ac.jp/c-elegans/blml/CE\_INDEX.html) and The C. elegans ORFome cloning project (http://worldb.dfci.harvard.edu/), similarity to other proteins from Blastx analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans Genbank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

#### FEATURES

##### source

1..30102

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db\_xref="taxon:6239"

/chromosome="I"

/clone="C32F10"

3494..3495

/note="SV1 trans-splice acceptor; see AF116529"

3516..11115

/gene="lin-35"

/note="for a graphical representation of this gene see: {www.wormbase.org/db/seq/sequence?name=C32F10.2;class=Sequ

##### misc\_feature

##### gene



ence)"  
join(3516. .3573,3826.3921,7275.7495,7758.7874,  
8133.8806,8857.9264,9413.9644,9693.10014,10075.10199,  
10483.11115)  
/gene="lin-35"  
/note="C. elegans synthetic multivulva protein LIN-35  
(GB:AF116529); coded for by the following C. elegans  
cDNAs: yk13c4.3, yk13c4.5, yk251e2.3, yk251e2.5,  
yk35a3.3, yk35a3.5, AF116529"  
/codon\_start=1  
/product="C. elegans LIN-35 protein (corresponding  
sequence C32F10.2)"  
/protein\_id="AAC24266.2"  
/db\_xref="GI:6671809"  
/translation="MKRAADPEGTSTDPFHEQSPFADVLAGETTDTTICEPPAKR  
IDLILKOEENGSGVSGGLIKNESELTQMTIKETEDINFEAREEEDBEDRSTSM  
PALLEDSDGFDGSLFDKNTPTSPQSFGCPAALPNVILGAEVYHNAVOR  
VSLSEESMOSAIYYLLYSKIRKRGITRILOPFPRICKITINSDISVAEMD  
KARVREHSKIRRYOYIRIOEGAGTIVYKFOECRITTEELIKVSGENPS  
SHELTVMTSFLVAKSRRTVDLISNOLLSILDQVTEMCSKEGIVYHNKNQFY  
EDLENDCTIRALCTOEGSVLDARHFSDFHFKMEKGTGIPSTNPOFE RDLINAVP  
KTYAENLQGRSIDIERTIPSEVDFSKTQSDPVSADILKVSYSGRFRALNTP  
KISNHNCEKLALAGKVASEKLVTOGSRPACVEYNELGYSFOLDSNNOLNLR  
LKIKSGMKEKLAEVCGTSDSPMAILKSDMTKNEFTTSLAEIGETINENIP  
KHYAVRKELEIVFLIFMEKTIIVAEIKKREEDILVIRBEEIDSVYFCCELTIV  
SKGYVRPFPMSAEICGVHFMFKHVIDLMTHEKOLSRMOWHFSIEEYVIEFSK  
SSSPMPHVRVRCPEFAHFOEGEDMAKLKSYSPKTPYIKRPDLEDELRIPVONQ  
TSRTLRIFLAKTYTFAARLODLTVBSNGARAKSQCSWLSFIDLNDPLIENDRIHD  
QILLCVFFIKINESMLFTEIMAVOROSANSLIYVSIVFQDLSDPENOAANT  
KTIIEERLEGPOKERTVDIIRKYNIIEFRDKRIYIIIGDISADEMDLMEPVATEGL  
MVRVYLFHKLISIQILPKTKHGESKORAIANLEKGIITIAERSGD"  
complement(4284. .5781)  
/gene="C32F10.4"  
/note="for a graphical representation of this gene see:  
[www.wormbase.org/db/seq/sequence?name=C32F10.4;class=Sequ  
ence]"  
complement(join(4284. .4424,4733.4964,5009.5183,  
5229.5322,5426.5605,5653.5781))  
/gene="C32F10.4"  
/note="coded for by the following C. elegans cDNAs:  
yk150f2.3, yk150f2.5"  
/codon\_start=1  
/product="Hypothetical protein C32F10.4"  
/protein\_id="AAC24267.1"  
/db\_xref="GI:1946999"  
/translation="MTTESSQTTDPYQHQRAEPHQPTORIAAEPRVFCGLPTIGG  
AGVIGISIMICQIFGLIYSPVLYVFEISGSAAMFVLSACKRKRTIATVAMYL  
LSMTFEMVLELIGAVTIACIVISDNNNDINPNOKDVTACKVCKRTIATVAMYL  
SCFELVTSIVAVLQKRYRMILLYNPLRRNAYRLPQOASNYNYQIVLPYPMSSGMP  
PRVYATAPMSSQTAPELPIYKQDPVDSSTTSASTSGSEFSTPIITVYANSEPGA  
DMSATINDPDRPRATSRPAYSVDVEEGHITVRSVDLQ"  
complement(11545. .13911)  
/gene="hmg-3"  
/note="for a graphical representation of this gene see:  
[www.wormbase.org/db/seq/sequence?name=C32F10.5;class=Sequ  
ence]"  
complement(join(11545. .11796,11935.12877,12966.13603,  
13675.13911))  
/gene="hmg-3"  
/note="Contains similarity to Pfam domain: PF00505  
(HMG\_box), Score=79.3, E-value=2.3e-20, N=1; coded for by  
the following C. elegans cDNAs: cm1h10, yk202g10.3,  
yk252d2.3, yk252d2.5, yk421h10.3, yk421h10.5"  
/codon\_start=1  
/product="C. elegans HMG-3 protein (corresponding sequence  
C32F10.5)"  
/protein\_id="AAC24268.1"  
/db\_xref="GI:1947000"  
/translation="MTELKFKGVYVEDIGHLTCGLTLTENSINFTIGDKGKSVYTG  
TVYDKIKQKOLNKPGLRVLGSDGAKHRGGFLDDDLQKOSTSSMKSINSQNF  
INGMNGADVAGKIKSEFMENPEIRFICITANSYIANKNEMLILFEHMOESKYOIM  
EKREHNPVLDNEEDTKYERKKAVLATAGLEAFETOPCLITDLTIFPRRDIK  
VPTISALHGKTYDKIIPVKTINRLVLPKHKDRQYVYVLSNMPYIRQGGTHSYLFE

Query Match	86.38; Score 16.4; DB 3; Length 30102;
Best Local Similarity	94.48; Pred. No. 2.7e+02;
Matches 1/; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Oy	1 CTGGGGCAGATTATGCA 18 
Db	5289 CTGGGACAGATTATGCA 5272
RESULT 13	
LOCUS	AF003145/38525 bp DNA linear INV 31-MAY-2002
DEFINITION	Caenorhabditis elegans cosmid B0414, complete sequence.
ACCESSION	AF003145
VERSION	AF003145.1 GI:2088760
KEYWORDS	HTG.
SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans.
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
AUTHORS	Rhabditioidea; Rhabditidae; Peloderae; Caenorhabditis.
TITLE	1 (bases 1 to 38525)
JOURNAL	Waterston, R.
PUBMED	Genome sequence of the nematode C. elegans: a platform for
REFERENCE	investigating biology. The C. elegans Sequencing Consortium
REFERENCE	Science 282 (5396), 2012-2018 (1998)
TITLE	2 (bases 1 to 38525)
JOURNAL	Sammons, L., Mohlmann, P. and Rohlfing, T.
REFERENCE	The sequence of C. elegans cosmid B0414
TITLE	3 (bases 1 to 38525)
JOURNAL	Unpublished (2001)
REFERENCE	Waterston, R.
TITLE	4 (bases 1 to 38525)
JOURNAL	Submitted (08-MAY-1997) Department of Genetics, Washington
REFERENCE	University, 4444 Forest Park Avenue, St. Louis, MO 63108, USA
AUTHORS	4 (bases 1 to 38525)
TITLE	Waterston, R.
JOURNAL	Submitted (05-JAN-2000) Department of Genetics, Washington
REFERENCE	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
TITLE	5 (bases 1 to 38525)
JOURNAL	Submitted (19-APR-2002) Department of Genetics, Washington
REFERENCE	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
AUTHORS	6 (bases 1 to 38525)
TITLE	Waterston, R.
JOURNAL	Submitted (19-APR-2002) Department of Genetics, Washington
REFERENCE	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
AUTHORS	Louis, MO 63110, USA
TITLE	7 (bases 1 to 38525)
JOURNAL	Submitted (31-MAY-2002) Department of Genetics, Washington
REFERENCE	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
AUTHORS	Louis, MO 63110, USA
TITLE	Submitted by:
JOURNAL	Genome Sequencing Center
REFERENCE	Department of Genetics, Washington University
AUTHORS	St. Louis, MO 63110, USA, and
TITLE	Sanger Centre, Hinxton Hall
JOURNAL	Cambridge CB10 0HQ, England
COMMENT	email: tw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this cosmid sequence and its analysis see:  
<http://www.wormbase.org/db/seq/sequence?name=B0414.3;class=Sequence>

## NEIGHBORING COSMID INFORMATION

The 5' cosmid is T22E7, 2000 bp overlap; the 3' cosmid is C32F10, 5900 bp overlap. Actual start of this cosmid is at base position 1 of B0414; actual end is at 7060 of C32F10.

## NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yujii Kohara ([http://www.ddbj.nig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html)) and The C. elegans ORFome cloning project (<http://worldb.dfci.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans Genbank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

## Location/Qualifiers

1. 38525

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db\_xref="taxon:6239"

/chromosome="I"

/clone="B0414"

925..1308

/gene="B0414.4"

/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=B0414.4;class=Sequence>"

join(925..1086,1147..1308)

/gene="B0414.4"

/codon\_start=1

/product="Hypothetical protein B0414.4"

/protein\_id="AAB57716.1"

/db\_xref="GI:2088763"

/translation="MNGSLVNTICQIPADIDPFLSSLSLISNNPITCCATRLTSSDDVFLMQLPDKASPELLEGGKRIFFELSDYCKTTRNPKKASFFQFLVLAIVGSEFLP"

3032..3868

/gene="B0414.3"

/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=B0414.3;class=Sequence>"

join(3032..3268,3428..3868)

/gene="B0414.3"

/note="Contains similarity to Pfam domain: PF00538 (linker-histone), Score=115.3, E-value=3.6e-31, N=1, coded for by the following C. elegans cDNAs: yk67e4.5, yk77a3.5, yk77a3.3, yk79e10.5, yk95h6.3, yk95h6.5, yk106g6.3, yk106g6.5, yk11a42.3, yk11a42.5, yk125e1.5, yk125e1.3, yk151a6.5, yk151a6.3, yk172d2.3, yk172d2.5, yk173g2.3, yk173g2.5, yk184c6.3, yk184c6.5, CEESX29F, CEMS55F,

CDS

gene

FEATURES

source

gene

CDS

CDS

gene

gene

CDS

gene

CDS

yk229f12.3, yk229f12.5, yk326d2.3, yk326d2.5, yk376a12.3, yk376a12.5, yk393a7.3, yk393a7.5, yk447f5.5, yk453f11.3, yk453f11.5, yk456d2.3, yk456d2.5, yk458f11.3, yk458f11.5"

/codon\_start=1

/product="Hypothetical protein B0414.3"

/protein\_id="AAB57720.1"

/db\_xref="GI:2088767"

/translation="MSDVVAEPPVAVKPTPKASKATKAKATKIPKVVAAAPPFLMTEAVSNLDRKSSRVAFETFKATKLTGDDVNTNALRSALKKGVSVKLYOTNGICANGFRVAABKBPAYKAKATGCKWKYTAKKAVASDRAKKYAKTKGCKYKVSFRKIAKPAVKRYTKAAAPTRSAABETAFKAAATAPKAAVTKATKTPARKA VETAPKA"

1252..1680

/gene="rnt-1"

/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=B0414.2;class=Sequence>"

join(1252..1320,13218..13335,13592..13711,14513..14586,15563..15679,16289..16347,16400..16518,16571..16800)

/gene="rnt-1"

/note="C. elegans homolog for Drosophila rnt (RNT-1); contains similarity to Pfam domain PF00853, Score=159.4, E-value=1.6e-44, N=1, coded for by the following C. elegans cDNAs: yk305f9.3, yk305f9.5, AB027412"

/codon\_start=1

/product="C. elegans RNT-1 protein (corresponding sequence B0414.2)"

/protein\_id="AAB57715.2"

/db\_xref="GI:6671807"

/translation="MTNFFHVRNFTIEQOPAPAKTLEKSSNLIYTLAPKMRSKSFEPEPVVLLTPVPNTVEVSIWAGNDEKCEEVREKAVHVOAKFNDLRFVSGRGRKFLITVTHSAPMMVAVKVIKVTGVPDAPRKPQSGLRQAEOQTLEPDDIRPQGPMPMTMTLPPMPFPLPMTQTPSPSPFPIISGPHSIALMKIIESEMKPIKQKEDVENLNTSSCSPSIFRTPSDQRKLKRRSPRSITKSETSIINTLQETPESVSKRRNRVNTSSSSSPITWRPE"

complement(17581..20134)

/gene="cpb-3"

/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=B0414.5;class=Sequence>"

complement(join(17581..17792,17868..18080,18157..18757,18814..18980,19035..19964,20012..20134))

/gene="cpb-3"

/note="strong similarity to Xenopus cytoplasmic polyadenylation element binding protein (NID:9624634); coded for by the following C. elegans cDNAs: yk88f12.3, yk88f12.5, yk220a12.3, yk220a12.5, yk240d6.5"

/codon\_start=1

/product="C. elegans CPB-3 protein (corresponding sequence B0414.5)"

/protein\_id="AAB57717.1"

/db\_xref="GI:2088764"

/translation="NNLNDRVIEIGENERSIVAKSTERRKKIPSLKLPSPVESGKSPTPAVYDLFFKKYKSDSVKPREEDMMNNGFDLSTDEKNGFLRKLQMLVGKSKSKYDESTSPANRLKLTVPKSRSTETHPNSNRHIGSSSTTKTSIDSENSORRKSARLNFEDAEQKDKSTKTNSSVFVAETRETVTSANESAPYIKRANRLIRNPKSLDLETPDSPOKGFSSSTESSPSDSMNQFPSEHRTSANGSPARKNFQDEHCKNDSQATWFGELPRLDYTPSPFKRIFVGGVPMIDTAAALKDSGEGSCAVEPEQEARYSGOSNLAPNSNGSLRAQTKYGAATGYVMI FEDERAVASLHECQSEIGAGAEVFXIRBORSKSTEIRVOGOPOYVOIIPWVTSQMSQEDDTLTLETIEPRKTVFVGAHLGHTMAYULSINEDCGAYECYQLODRKRYPIGSGRYFRHGCALDGLKDDQDVHLPVIRKRVQIDPLESTCCWCNSQSAHCFCKRNLCKFYKCHCALDGLKDDQDVHLPVIVSSASAKFPGATTRRSQVSSLSLKNSNNSQVNSLPHEPSAGFPIIVGAPQATLSALYGIQNGOQLKPAVEFPWTPSPSEINRRSRFTFSPPTVFPSTPVSIPKSVSDGSLPAAVYNSAFLPPLPYGSPNHPSTSNISQOQPYGANTLYGFIPOQHP SAMRSPVNYGO"

complement(20992..23452)

/gene="g1h-3"

/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=B0414.6;class=Sequence>"

Query Match 86.3%; Score 16.4; DB 3; Length 38525;  
Best Local Similarity 94.4%; Pred. No. 2.7e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

0Y 1 CTGGGGGCGAGTTATTGCA 18  
Db 37914 CTGGGAGCAGTTATTGCA 37897  
||||| |||||||

RESULT 14  
LOCUS U00015 42325 bp DNA linear BCT 01-MAR-1994  
DEFINITION Mycobacterium leprae cosmid B1620.  
ACCESSION U00015  
VERSION U00015.1 GI:466931  
KEYWORDS  
SOURCE Mycobacterium leprae.  
ORGANISM Mycobacterium leprae  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
Mycobacterium.  
REFERENCE 1 (bases 1 to 42325)  
AUTHORS Smith,D.R.  
JOURNAL Unpublished  
REFERENCE 2 (bases 38281 to 41190)  
AUTHORS Rinke de Wit,T.F., Bekelie,S., Osland,A., Miko,T.L., Hermans,P.W.,  
van Soelingen,D., Drijfhout,J.W., Schonlingh,R., Janson,A.A. and  
Thole,J.E.  
Mycobacteria contain two groEL genes: the second Mycobacterium  
leprae groEL gene is arranged in an operon with groES  
MoI. Microbiol. 6 (14), 1995-2007 (1992)  
JOURNAL 92374850  
MEDLINE 1354834  
PUBMED  
REFERENCE 3 (bases 3688 to 5472)  
AUTHORS Vega-Lopez,F., Brooks,L.A., Dockrell,H.M., De Smet,K.A.,  
Thompson,J.K., Hussain,R. and Stoker,N.G.  
Sequence and immunological characterization of a serine-rich  
antigen from Mycobacterium leprae  
Infect. Immun. 61 (5), 2145-2153 (1993)  
JOURNAL 93239328  
MEDLINE 8478104  
PUBMED  
REFERENCE 4 (bases 2759 to 5248)  
AUTHORS Rinke de Wit,T.F., Clark-Curtiss,J.E., Abebe,F., Kolk,A.H.,  
Janson,A.A., van Agterveld,M. and Thole,J.E.  
A Mycobacterium leprae-specific gene encoding an immunologically  
recognized 45 kDa protein  
MoI. Microbiol. 10 (4), 829-838 (1993)  
JOURNAL 95020554  
MEDLINE 7934845  
PUBMED  
REFERENCE 5 (bases 1 to 42325)  
AUTHORS Robison,K.  
JOURNAL Direct Submission  
TITLE Submitted (01-NOV-1993) Department of Genetics, Harvard Medical  
School, 200 Longwood Avenue, Boston MA 02115  
6 (bases 1 to 42325)  
Robison,K.  
Direct Submission  
Submitted (01-MAR-1994) Department of Genetics, Harvard Medical  
School, 200 Longwood Avenue, Boston MA 02115  
On Mar 31, 1994 this sequence version replaced gi:414221.  
This sequence data was produced by the genome sequencing center  
located at Collaborative Research Incorporated (1365 Main St.,  
Waltham MA, 02159). 617-487-7979. Please contact Doug Smith  
(smith@cric.com). The annotation should be considered  
preliminary and incomplete.  
Location/Qualifiers  
1..42325  
/organism="Mycobacterium leprae"  
/db\_xref="taxon:1769"  
619..1908  
/EC\_number="3.1.3.12"  
/note="Trehalose-6-phosphate phosphatase; B1620\_F1\_1"  
/codon\_start=1

/transl\_table=11  
/product="otsp"  
/protein\_id="AAC43238.1"  
/db\_xref="GI:466950"  
/translation="MPVTIDPRHSASLFIDAVVDTPTPLDSTVTYLVKQLOGICVGA  
VESTRNSQGVLTATGDLHPVAVDGLASKAVITLVAANARLMAQPCRCVAVVADA  
GITAARYGCFALVGLVIGEDQGTGRDLRNSGADTVADAGEIVTRGDRMSLELP  
ALQITLGLTDLTAROPAVFEPDEGTLSDIYDDPSARVPQATFALQKATRCVAIL  
SGRLADYIKRIGVPGIWMYSGSHGFEAPDGTGHONDAEATPIEEOATQIRDOI  
GPPIGVAYEHRKRFQVAVYRVARVDREAVAVATAGQARALVYTTGRIEELRPDI  
DWDKGTLHWVIDRLRHAGTQVGSASLMPICLDGDDTDEDAFDVARRHDTVGIGIYVVR  
HTEGDNRAFLALFTLDSPMHVSFEETRLARQLSDTOR"  
complement(1979..2188)  
/note="weak match 42.1 kD protein; B1620\_C3\_270"  
/codon\_start=1  
/transl\_table=11  
/product="u1620m"  
/protein\_id="AAC43237.1"  
/db\_xref="GI:466949"  
/translation="MVCIRGLAGFLKPILYLSRYGVPRALVVELKGLAPNLSIT  
ITGVSRPADHLDLENRARDGVSY"  
complement(2282..2473)  
/note="weak match to (sp:x66077); B1620\_C3\_269"  
/codon\_start=1  
/transl\_table=11  
/product="u1620n"  
/protein\_id="AAC43236.1"  
/db\_xref="GI:466948"  
/translation="MDHRSLSCRDKEQEPDRCVMPVRYGPIWDAESLIRPRMS  
SRVWGSVATAGTGFCGS"  
complement(3910..5136)  
/note="45 kDa serine-rich antigen; B1620\_C1\_177"  
/codon\_start=1  
/transl\_table=11  
/product="g45"  
/protein\_id="AAC43220.1"  
/db\_xref="GI:466932"  
/translation="MEDFMYSPVEYNAFLMRSQSPPLMGAAEWISLABOLMEAAQ  
EVSPTIYVAVPASFAGERSDMLASRSVFTVWIDGMARNAGLIRVIAHAVYAEER  
AGMPPLTLVLTNIITHTALKAIRNFQGVSTTVAALAEADYDLMTQNSTAMTYTRDTVL  
RBTGKMEYFEPAPOLVSRKCMRDSDVSFRSSSDSLYESIDNLDVSAQSEHGSL  
DMSQSVNTGCSVAQSELCDSPFGTPSQSSQNDLSAVSLQQLGDLSDIISASASL  
LTTNLSISSSTASIMPIVASQVTEITIGSOYAVEKMQSISTFASVDAASKVAVG  
GOAVSYGLARVPENMATSPVMAFASHPAGCAITTAVSGPLEGVYQPAEVLITAS  
VAGSGTCGAPAFNAV"  
complement(5262..5564)  
/note="weak match 32kD protein (sp:p16645); B1620\_C2\_218"  
/codon\_start=1  
/transl\_table=11  
/product="u1620k"  
/protein\_id="AAC43225.1"  
/db\_xref="GI:466937"  
/translation="WSEFLRYEVGGLMAAGRLERTSESACNAKLTPTTKVPPA  
ADQVSKLVQVFSYSGKQEGYAAQGVQDSRLFYQSLKDAAGDYWDSHMYLNTD"  
6340..6816  
/note="transposase; B1620\_F1\_4"  
/codon\_start=1  
/transl\_table=11  
/product="u1620a"  
/protein\_id="AAC43241.1"  
/db\_xref="GI:466953"  
/translation="MEDICVANNRYGVKRLNRTAKRAGHDHRDHQIDGIVAVKCLAG  
LDGIVDRKRIITRETHEDQAPGRHVRPQCLVAKYGVSVESPIEMPRACVAG  
TAFQVDVFSKRIIPGPGRTSTTYLVLSLFDALFIRKTYELREFPADFSAPLQGR"  
complement(7286..7606)  
/note="transposase; B1620\_C3\_266"  
/codon\_start=1  
/transl\_table=11  
/product="u1620b"  
/protein\_id="AAC43235.1"  
/db\_xref="GI:466947"  
/translation="MVNVHALITTDVNAEGYRETLGCGCDHRRRGRGLVEVSAVADRP  
LAFRGLYGHQRCPRRTARRHNOGQPPGVFMQRCRIYLGHPHGLATHGFLVNGAVPAT

```

CDS
    ILGTV"
    complement(8447..9097)
    /codon_start=1
    /transl_table=11
    /product="B1620_C3_264"
    /protein_id="AAC43234.1"
    /db_xref="GI:466946"
    /translation="MGRIRIRRLPMPSTFPAQOEFRDVVVVEPGNDPDRDCDTV
    DMVGGAHMTIDPVGFASSAPQDRGTVDNTRGAEKEKPEVAFTYTNR
    PRTVSVYMDGRIDHVELSKRTWSEGVSELIYADLAKRKAQSOYFIIDKL
    SOLADGDEHRYALRESVGNNTNLPSEQAALAEVATRTSDICPADTENDOM"
    complement(9111..9431)
    /codon_start=1
    /transl_table=11
    /product="B1620_C2_214"
    /protein_id="AAC43224.1"
    /db_xref="GI:466936"
    /translation="MRKMDNTVQSEHNSIASQHENENACASSGVSAAGLANVAS
    TSHGSCAOFNDTLNMYEDAHHTLDESLHTGGIDILARVLRAAAMYCDADETCGSDIK
    SAFC"
    complement(9489..10673)
    /codon_start=1
    /transl_table=11
    /product="B1620_C2_213"
    /protein_id="AAC43223.1"
    /db_xref="GI:466935"
    /translation="MSGAFIIDPTLKATIEAMHALLIGVNDGVYLSLSLFFERALE
    HLAAPGPGMGSAADKACONRRKVDIFQELAEDELKELILHNOANSYOTGIL
    DKAKKALFEVPAIDNLTIPVGSVMSASIOACAAAMAASGSLAFLVOTAIHT
    AKFVALLARLALHASAVADYVDCVAIIKGIYDHLMHFIAGALGIDIVKRIIHF
    RLFESHMSRHSFRFGIPGLSGATSGLSQVYGLGVPLAASSGSLSESLSTENT
    PSLAGVAGLGLSLPQLAQHLAASTRQSTROAGVSAELSTDFGQGOEPVSAQSO
    GNGSGQMGGMTPASTSKSKDKRRKKYSEGAAGTDAERAPIEVQSGGRALAOH
    VV"
    13956
    /codon_start=1
    /transl_table=11
    /product="B1620_F1_14"
    /protein_id="AAC43339.1"
    /db_xref="GI:466951"
    /translation="MSKPLQPECADHIAAVLFGMTNFOAMNFGESPVAIGPAHQAK
    AKVGVKPRCALNTSISLPTVTQTTSAWPRREQLRPIALS"
    16793..17830
    /note="D-Ribose periplasmic binding protein; B1620_F2_67"
    /codon_start=1
    /transl_table=11
    /product="rdsb"
    /protein_id="AAC43242.1"
    /db_xref="GI:466954"
    /translation="MSTGTSVPLRVASIMLIIGLLPAILPACGSESHHKTITISYP
    TANSFFMAVYRFRIDEGSHQDGVYINAVSAEEDQKQSDVETLLNQGIDGLIVPOS
    TSIAPTLRIATQANIPIVVYVDRPYAPGQKNADYVAFLEPNNEKSGSJAELMA
    GGSKEFLAIGMGNSVAGKAGLESALTAVGHRLVOFQVQNSEKSLGAENMLA
    AHPADGANAIFCENDKICOGAIIKAVYNNREKEPIEGGMDLIPQALAIENGLYVSL
    GAHMLGEGGGLAILRKIHGDPALRMKYLDLKYDKSVAKFYRIDNTRHYMKS
    TTPEDMTITLN"
    18218..19951
    /codon_start=1
    /transl_table=11
    /product="B1620_F2_68"
    /protein_id="AAC43243.1"
    /db_xref="GI:466955"
    /translation="MWDIDGRLESINPATRITLGRANDVDMKHGHPFCFYDTONR
    VDLEREVMVREVTYSKYVIGDQSGRLMSLVNSILAKAPPHSAIVSFSDI
    SAHMLIERLTAEATHDCLGLANRRPAEDQITKSLQDERSRLAAVILLDDPRKVI
    
```

Query Match 86.3%; Score 16.4; DB 1; Length 42325;  
 Best Local Similarity 94.4%; Pred. No. 2.7e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTGGGGGCGAGTATTGCA 18  
 ||||| ||||| ||||| |||||  
 DB 29336 CTGGAGGCGAGTATTGCA 29319

RESULT 15  
 AC101474  
 LOCUS  
 DEFINITION  
 AC101474  
 VERSION  
 AC101474.1 GI:17060249  
 KEYWORDS  
 HTG: HTGS\_PHASE0.  
 SOURCE  
 Mus musculus.  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 65707)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Mus musculus, clone RP23-187L16  
 Unpublished  
 2 (bases 1 to 65707)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,  
 Anderson,S., Barin,N., Bastien,V., Boguslavsky,L., Boukhallier,B.,  
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
 Choquel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,  
 Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
 Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Garbino,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hages,B., Heathord,A., Horton,L., Hulme,W., Illey,I., Johnson,R.,  
 Jones,C., Kamat,A., Karatas,A., Kells,C., Laroque,K.,  
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
 Maclean,C., MacDonald,P., Major,J., Margulis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,  
 Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
 Norou,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green,P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE  
 JOURNAL  
 COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: L16819  
 Center clone name: 187\_L16  
 \*\*\*\*\*  
 \* NOTE: This record contains 82 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.  
 \* 1  
 \* 725 824: gap of 100 bp  
 \* 825 1548: contig of 724 bp in length  
 \* 1549 1648: gap of 100 bp  
 \* 1649 2372: contig of 724 bp in length  
 \* 2373 2472: gap of 100 bp  
 \* 2473 3152: contig of 680 bp in length

3153 3252: gap of 100 bp  
3253 3923: contig of 671 bp in length  
3924 4023: gap of 100 bp  
4024 4694: contig of 671 bp in length  
4695 4794: gap of 100 bp  
4795 5497: contig of 703 bp in length  
5498 5597: gap of 100 bp  
5598 6307: contig of 710 bp in length  
6308 6407: gap of 100 bp  
6408 7120: contig of 713 bp in length  
7121 7220: gap of 100 bp  
7221 7935: contig of 715 bp in length  
7936 8035: gap of 100 bp  
8036 8760: contig of 725 bp in length  
8761 8860: gap of 100 bp  
8861 9589: contig of 729 bp in length  
9590 9689: gap of 100 bp  
9690 10407: contig of 718 bp in length  
10408 10507: gap of 100 bp  
10508 11190: contig of 683 bp in length  
11191 11290: gap of 100 bp  
11291 11966: contig of 676 bp in length  
11967 12066: gap of 100 bp  
12067 12763: contig of 697 bp in length  
12764 12863: gap of 100 bp  
12864 13576: contig of 713 bp in length  
13577 13676: gap of 100 bp  
13677 14385: contig of 709 bp in length  
14386 14485: gap of 100 bp  
14486 15193: contig of 708 bp in length  
15194 15293: gap of 100 bp  
15294 16002: contig of 709 bp in length  
16003 16102: gap of 100 bp  
16103 16820: contig of 718 bp in length  
16821 16920: gap of 100 bp  
16921 17606: contig of 686 bp in length  
17607 17706: gap of 100 bp  
17707 18431: contig of 725 bp in length  
18432 18531: gap of 100 bp  
18532 19258: contig of 727 bp in length  
19259 19358: gap of 100 bp  
19359 20048: contig of 690 bp in length  
20049 20148: gap of 100 bp  
20149 20838: contig of 690 bp in length  
20839 20938: gap of 100 bp  
20939 21605: contig of 667 bp in length  
21606 21705: gap of 100 bp  
21706 22405: contig of 700 bp in length  
22406 22505: gap of 100 bp  
22506 23202: contig of 697 bp in length  
23203 23302: gap of 100 bp  
23303 24003: contig of 701 bp in length  
24004 24103: gap of 100 bp  
24104 24816: contig of 713 bp in length  
24817 24916: gap of 100 bp  
24917 25611: contig of 695 bp in length  
25612 25711: gap of 100 bp  
25712 26440: contig of 729 bp in length  
26441 26540: gap of 100 bp  
26541 27232: contig of 692 bp in length  
27233 27332: gap of 100 bp  
27333 27980: contig of 648 bp in length  
27981 28080: gap of 100 bp  
28081 28746: contig of 666 bp in length  
28747 28846: gap of 100 bp  
28847 29552: contig of 706 bp in length  
29553 29652: gap of 100 bp  
29653 30358: contig of 706 bp in length  
30359 30458: gap of 100 bp  
30459 31159: contig of 701 bp in length  
31160 31259: gap of 100 bp  
31260 31974: contig of 715 bp in length  
31975 32074: gap of 100 bp

32075 32804: contig of 730 bp in length  
32805 32904: gap of 100 bp  
32905 33578: contig of 674 bp in length  
33579 33678: gap of 100 bp  
33679 34377: contig of 699 bp in length  
34378 34477: gap of 100 bp  
34478 35204: contig of 727 bp in length  
35205 35304: gap of 100 bp  
35305 35986: contig of 682 bp in length  
35987 36086: gap of 100 bp  
36087 36759: contig of 673 bp in length  
36760 36859: gap of 100 bp  
36860 37527: contig of 668 bp in length  
37528 37627: gap of 100 bp  
37628 38331: contig of 704 bp in length  
38332 38431: gap of 100 bp  
38432 39144: contig of 713 bp in length  
39145 39244: gap of 100 bp  
39245 39950: contig of 706 bp in length  
39951 40050: gap of 100 bp  
40051 40776: contig of 726 bp in length  
40777 40876: gap of 100 bp  
40877 41590: contig of 714 bp in length  
41591 41690: gap of 100 bp  
41691 42408: contig of 718 bp in length  
42409 42508: gap of 100 bp  
42509 43186: contig of 678 bp in length  
43187 43286: gap of 100 bp  
43287 43949: contig of 663 bp in length  
43950 44049: gap of 100 bp  
44050 44750: contig of 701 bp in length  
44751 44850: gap of 100 bp  
44851 45550: contig of 700 bp in length  
45551 45650: gap of 100 bp  
45651 46325: contig of 675 bp in length  
46326 46425: gap of 100 bp  
46426 47139: contig of 714 bp in length  
47140 47239: gap of 100 bp  
47240 47963: contig of 724 bp in length  
47964 48063: gap of 100 bp  
48064 48786: contig of 723 bp in length  
48787 48886: gap of 100 bp  
48887 49612: contig of 726 bp in length  
49613 49712: gap of 100 bp  
49713 50434: contig of 722 bp in length  
50435 50534: gap of 100 bp  
50535 51223: contig of 689 bp in length  
51224 51323: gap of 100 bp  
51324 51998: contig of 675 bp in length  
51999 52098: gap of 100 bp  
52099 52776: contig of 678 bp in length  
52777 52876: gap of 100 bp  
52877 53580: contig of 704 bp in length  
53581 53680: gap of 100 bp  
53681 54382: contig of 702 bp in length  
54383 54482: gap of 100 bp  
54483 55186: contig of 704 bp in length  
55187 55286: gap of 100 bp

Query Match 86.3%; Score 16.4; DB 2; Length 65707;  
Best Local Similarity 94.4%; Pred. No. 2.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTGGGGCAGTTATTCGA 18

DB 34780 CTGGGAGCAGTTATTCGA 34797

Search completed: July 1, 2003, 07:23:04  
J00 time : 363.359 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 03:06:29 ; Search time 105.231 Seconds  
(without alignments)  
406.611 Million cell updates/sec

Title: US-10-053-662A-30

Perfect score: 19

Sequence: 1 ctg999gcagttatgcac 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_101002:\*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match	Length	DB	ID	Description
1	16.4	86.3	271	14	AA060904	Human brain Expres
2	15.8	83.2	1095	22	AAS29516	Human endocrine po
3	15.8	83.2	1183	22	ABA06384	Human cDNA SEQ ID
4	15.8	83.2	1183	22	AAS29640	Human endocrine po
5	15.8	83.2	1183	22	AA163963	Human polynucleoti
6	15.8	83.2	1183	22	AAS31277	Human cDNA encodin
7	15.8	83.2	1183	22	AAS34773	cDNA encoding nove
8	15.8	83.2	1183	24	ABQ6601	Human polynucleoti
9	15.8	83.2	1490	20	AA252932	Human prostate tum

#### ALIGNMENTS

RESULT 1	AA060904	standard; DNA; 271 BP.
AA060904:		
AC	AA060904:	
XX		
DT	16-MAR-1994 (first entry)	
XX		
DE	Human brain Expressed Sequence Tag EST000995.	
XX		
KW	Gene transcription product; genetic markers; tagging; in vivo;	
KW	transcription; mapping; locations; chromosomes; chromosomal; ss.	
OS	Homo sapiens.	
XX		
PN	W0916178-A.	
XX		
PD	19-AUG-1993.	
XX		
PF	12-FEB-1993;	93WO-US01294.
XX		
PR	12-FEB-1992;	92US-0837195.
XX		
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICE.	
XX		
PI	Adams MD, Moreno RF, Venter CJ;	
XX		
DR	WPI, 1993-272882/34.	
XX		
PT	Enriched oligonucleotides and corresp. sequences - used as	
PT	markers for human genes transcribed in-vivo, facilitate tagging	
PT	of most human genes	

C 10	15.8	83.2	2613	22	AA163904	Human polynucleoti
C 11	15.8	83.2	3463	21	AAC77382	Human OREX ORE2937
C 12	15.8	83.2	3531	23	AA568056	DNA encoding novel
C 13	15.8	83.2	6230	22	AA542520	Human cDNA encodin
C 14	15.4	81.1	164	23	ABV12689	Human prostate exp
C 15	15.4	81.1	427	23	ABV33819	Human prostate exp
C 16	15.4	81.1	3582	24	ABV42711	Human prostate exp
C 17	15.4	81.1	3620	21	AAD31145	Human laminin 5 cd
C 18	15.4	81.1	3720	21	AAC83732	Human laminin 5 cd
C 19	15.4	81.1	4316	17	AAT13324	Kallinin/laminin 5
C 20	15.4	81.1	4316	24	AAL42911	Laminin gamma-2 ch
C 21	15.4	81.1	5020	21	AAC83730	Human laminin 5 cd
C 22	15.4	81.1	5152	20	AA224627	Human lung tumor a
C 23	15.4	81.1	5156	21	AA658866	Human lung cancer-
C 24	15.4	81.1	5156	24	AB149085	Human lung tumour
C 25	15.4	81.1	5200	17	AAT13323	Kallinin/laminin 5
C 26	15.4	81.1	5200	21	AAC83729	Human laminin 5 cd
C 27	15.4	81.1	5200	21	AAL42910	Laminin gamma-2 ch
C 28	15.4	81.1	5200	23	ABK42922	Genomic sequence #
C 29	15.4	81.1	11443	19	AAV52182	Streptococcus pneu
C 30	14.8	77.9	403	22	AA185494	Human polynucleoti
C 31	14.8	77.9	471	22	AA114658	Probe #4591 for ge
C 32	14.8	77.9	471	22	AA104468	Probe #4459 used t
C 33	14.8	77.9	475	24	AB140436	Sunflower lipid tr
C 34	14.8	77.9	566	22	ABA63470	Human foetal liver.
C 35	14.8	77.9	566	22	ABA30668	Probe #9134 for ge
C 36	14.8	77.9	566	22	AAK12003	Human brain expres
C 37	14.8	77.9	566	22	AAK37709	Human bone marrow
C 38	14.8	77.9	566	22	AA118465	Probe #8398 for ge
C 39	14.8	77.9	566	22	AA135882	Human genome-deriv
C 40	14.8	77.9	566	24	AB116599	Human musculoskele
C 41	14.8	77.9	694	22	AA137202	Human musculoskele
C 42	14.8	77.9	694	22	AA137204	Streptococcus pneu
C 43	14.8	77.9	867	21	AA05387	Aspergillus oryzae
C 44	14.8	77.9	867	21	AA05387	
C 45	14.8	77.9	1213	21	AA13700	

XX Example 4; Page 409; 500bp; English.  
PS The Expressed Sequence Tag was isolated from a human brain cDNA  
XX library as part of a large set of ESTs which can be used as markers  
CC for human genes transcribed in vivo. They can be used to facilitate  
CC tagging of most human genes, for mapping locations of expressed genes  
CC on chromosomes, for individual or forensic identification, for mapping  
CC locations of disease-associated genes, for identification of tissue  
CC type, and for prep. of antisense sequences, probes and constructs.  
CC EST00995 has a "poor" coding probability as evaluated using the  
CC coding-region prediction program CRM. See also AA039041-Q01440.  
XX  
SQ Sequence 271 BP; 54 A; 98 C; 57 G; 60 T; 2 other;  
Query Match 86.3%; Score 16.4; DB 14; Length 271;  
Best Local Similarity 89.5%; Pred. No. 55;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CTGGGGGCGAGTTATTCGAC 19  
Db 218 CTGGGGGCGTGTATTTCAC 236  
||||||| ||||| |||  
RESULT 2  
AAS29516/c AAS29516 standard; cDNA; 1095 BP.  
XX  
AC AAS29516;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Human endocrine polypeptide encoding cDNA SEQ ID NO 16.  
XX  
XX  
KW cat; dog; chicken; sheep; immunosuppressive; antiarthritic; vasotropic;  
KW antirheumatic; antiproliferative; cytostatic; cardiac; neuroprotective;  
KW cerebroprotective; noctropis; antibacterial; vitucide; fungicide; cancer;  
KW ophthalmological; vulnerrary; gene therapy; autoimmune disease; neoplasm;  
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;  
KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
KW gastrointestinal disorder; renal disorder; respiratory disorder;  
KW wound healing; skin aging; organ transplantation; food preservative;  
KW tissue regeneration; anti-fertility.  
XX  
XX Homo sapiens.  
OS  
XX  
PN MO200155364-A2.  
PD  
XX 02-AUG-2001.  
PF  
XX 17-JAN-2001; 2001WO-US01308.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0226279.  
PR 18-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234423.  
PR 21-SEP-2000; 2000US-0234424.  
PR 25-SEP-2000; 2000US-0234497.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.



PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-451936/48.  
XX P-PSDB: AAU18287.  
XX  
XX Isolated polypeptide for treating, preventing and/or prognosing  
XX disorders of the endocrine system such as reproductive disorders,  
XX endocrine cancers and also for testing and detection e.g. diagnosis -  
PS  
PS Claim 1; SEQ ID NO 16; 604pp; English.  
XX  
XX Sequences AAS29511-AAS29736 represent cDNA molecules, which encode the  
XX endocrine polypeptides of the invention. Endocrine polypeptides and their  
XX associated polypeptides of the invention are useful in the diagnosis,  
XX treatment and prevention of various types of disorders in e.g. humans,  
XX mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A  
XX pathological condition can be determined by determining the presence or  
XX absence of a mutation in an endocrine polynucleotide. The treatable  
XX disorders include autoimmune diseases such as rheumatoid arthritis,  
XX hyperproliferative disorders such as neoplasms of the breast or liver,  
XX cardiovascular disorders such as cardiac arrest, cerebrovascular  
XX disorders such as cerebral ischaemia, nervous system disorders such as  
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi,  
XX ocular disorders such as corneal infection, endocrine disorders such as  
XX premature labour and infertility, gastrointestinal disorders such as  
XX Crohn's disease, renal disorders such as glomerulonephritis and

CC respiratory disorders such as asthma. The polypeptides can also be used  
CC to aid wound healing, to prevent skin aging due to sunburn, to maintain  
CC organs before transplantation, to regenerate tissues and in chemotaxis.  
CC The polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO

Query Match 83.2%; Score 15.8; DB 22; Length 1095;  
Best Local Similarity 89.5%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGCAGTATTCAC 19  
||||||| ||||| |||  
DB 663 CTGGGGCCTGTTATTCAC 645

RESULT 3  
ABA06384/C  
ID ABA06384 standard; cDNA; 1183 BP.  
XX  
XX ABA06384;  
XX  
XX 10-JAN-2002 (first entry)  
XX  
XX Human cDNA SEQ ID NO: 50.  
DE  
XX  
XX Human; gene therapy; neural disorder; immune system disorder;  
XX muscular disorder; reproductive disorder; gastrointestinal disorder;  
XX pulmonary disorder; cardiovascular disorder; renal disorder;  
XX proliferative disorder; inflammation; ss.  
OS  
XX Homo sapiens.  
PN  
XX WO200154474-A2.  
PD  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01349.  
XX  
XX 31-JAN-2000; 2000US-179065P.  
XX 04-FEB-2000; 2000US-180628P.  
XX 24-FEB-2000; 2000US-184664P.  
XX 02-MAR-2000; 2000US-186350P.  
XX 16-MAR-2000; 2000US-189874P.  
XX 17-MAR-2000; 2000US-190076P.  
XX 18-APR-2000; 2000US-198123P.  
XX 19-MAY-2000; 2000US-205515P.  
XX 07-JUN-2000; 2000US-209467P.  
XX 28-JUN-2000; 2000US-214866P.  
XX 30-JUN-2000; 2000US-215135P.  
XX 07-JUL-2000; 2000US-216647P.  
XX 07-JUL-2000; 2000US-217487P.  
XX 11-JUL-2000; 2000US-217496P.  
XX 14-JUL-2000; 2000US-218290P.  
XX 26-JUL-2000; 2000US-220963P.  
XX 26-JUL-2000; 2000US-220964P.  
XX 14-AUG-2000; 2000US-224518P.  
XX 14-AUG-2000; 2000US-224519P.  
XX 14-AUG-2000; 2000US-225213P.  
XX 14-AUG-2000; 2000US-225214P.  
XX 14-AUG-2000; 2000US-225266P.  
XX 14-AUG-2000; 2000US-225267P.  
XX 14-AUG-2000; 2000US-225268P.  
XX 14-AUG-2000; 2000US-225270P.  
XX 14-AUG-2000; 2000US-225447P.  
XX 14-AUG-2000; 2000US-225757P.  
XX 14-AUG-2000; 2000US-225758P.  
XX 14-AUG-2000; 2000US-225759P.  
XX 18-AUG-2000; 2000US-226279P.  
XX 22-AUG-2000; 2000US-226681P.  
XX 22-AUG-2000; 2000US-226686P.



KW antineumatic; antiproliferative; cytostatic; cardiant; neuroprotective;  
KW cerebroprotective; nootropic; antibacterial; virocidic; fungicide; cancer;  
KW ophthalmologic; vulnerary; gene therapy; autoimmune disease; neoplasm;  
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;  
KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
KW gastrointestinal disorder; renal disorder; respiratory disorder;  
KW wound healing; skin aging; organ transplantation; food preservative;  
KW tissue regeneration; anti-infertility.  
XX  
OS Homo sapiens.  
XX  
PN WO200155364-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01308.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUL-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 26-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234023.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236337.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249329.  
PR 17-NOV-2000; 2000US-0249330.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.

PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-451936/48.  
XX P-PSDB; AAU18411.  
XX  
XX Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders of the endocrine system such as reproductive disorders,  
PT endocrine cancers and also for testing and detection e.g. diagnosis -  
XX  
XX Claim 1; SEQ ID No 140; 604pp; English.  
XX  
XX Sequences AAS29511-AAS29736 represent cDNA molecules, which encode the  
CC endocrine polypeptides of the invention. Endocrine polypeptides and their  
CC associated polynucleotides of the invention are useful in the diagnosis,  
CC treatment and prevention of various types of disorders in e.g. humans,  
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A  
CC pathological condition can be determined by determining the presence or  
CC absence of a mutation in an endocrine polynucleotide. The treatable  
CC disorders include autoimmune diseases such as rheumatoid arthritis,  
CC hyperproliferative disorders such as neoplasms of the breast or liver,  
CC cardiovascular disorders such as cardiac arrest, cerebrovascular  
CC disorders such as cerebral ischemia, nervous system disorders such as  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,  
CC ocular disorders such as corneal infection, endocrine disorders such as  
CC premature labour and infertility, gastrointestinal disorders such as  
CC Crohn's disease, renal disorders such as glomerulonephritis and  
CC respiratory disorders such as asthma. The polypeptides can also be used  
CC to aid wound healing, to prevent skin aging due to sunburn, to maintain  
CC organs before transplantation, to regenerate tissues and in chemotaxis.  
CC The polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO

Query Match 83.2%; Score 15.8; DB 22; Length 1183;  
Best Local Similarity 89.5%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGGCGATTATTCAC 19  
DB 743 CTGGGGGCGATTATTCAC 725

RESULT 5  
ID AAI63963 standard; cDNA; 1183 BP.  
XX  
XX AAI63963;  
XX  
XX 22-OCT-2001 (first entry)  
XX  
XX Human polynucleotide SEQ ID No 171.  
XX  
XX Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;  
XX cerebroprotective; nootropic; neuroprotective; antibacterial; virocidic;  
XX fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;  
XX neuroprotective; antiallergic; hepatotropic; antidiabetic;  
XX antiinflammatory; anticancer; anticonvulsant; antibacterial;  
XX antiparasitic; cardiac; gene therapy; cancer; immune disorder;  
XX cardiovascular disorder; neurological disease; infection; human; ss.  
OS Homo sapiens.

XX  
XX WO200155308-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01309.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226868.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231968.  
XX 14-SEP-2000; 2000US-0232397.  
XX 14-SEP-2000; 2000US-0232398.  
XX 14-SEP-2000; 2000US-0232399.  
XX 14-SEP-2000; 2000US-0232400.  
XX 14-SEP-2000; 2000US-0232401.  
XX 14-SEP-2000; 2000US-0233063.  
XX 14-SEP-2000; 2000US-0233064.  
XX 14-SEP-2000; 2000US-0233065.  
XX 21-SEP-2000; 2000US-0234223.  
XX 21-SEP-2000; 2000US-0234274.  
XX 25-SEP-2000; 2000US-0234997.  
XX 25-SEP-2000; 2000US-0234998.  
XX 26-SEP-2000; 2000US-0235484.  
XX 27-SEP-2000; 2000US-0235834.

```
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239933.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0255678.
XX
XX
PA (HUMA - ) HUMAN GENOME SCI INC.
XX
```

```
PI Rosen CA, Barash SC, Ruben SM;
XX
XX MPI: 2001-488781/53.
DR
DR P-PSDB; AAM43657.
XX
XX New isolated nucleic acids and polypeptides, useful for diagnosing,
PI treating and/or preventing human diseases and disorders -
XX
XX Claim 1; SEQ ID NO 171; 664pp + Sequence listing; English.
XX
XX The invention relates to human polynucleotides (AA163803-AA164012) and
CC the encoded proteins (AAM434497-AAM43660) useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. The
CC genes were isolated from a range of human tissues disclosed in the
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 1183 BP; 299 A; 272 C; 363 G; 247 T; 2 other;
SQ
```

```
Query Match 83.2%; Score 15.8; DB 22; Length 1183;
Best Local Similarity 89.5%; Pred. NO.1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CTGGGGCGCATTATTCAC 19
DB 743 CTGGGGCGCATTATTCAC 725
```

```
RESULT 6
AAS31277/c
ID AAS31277 standard; cDNA; 1183 BP.
XX
XX AAS31277;
AC
AC
XX
XX 04-DEC-2001 (first entry)
DT
XX
XX Human cDNA encoding a novel extracellular matrix protein, Seq ID No 91.
DE
XX
XX Human; secreted extracellular matrix protein; ss; immunomodulatory;
KW Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cyostatic;
KW antiA2heimers; immune/autoimmune disease; HIV infection; anaemia;
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW cancer; hyperproliferative disorder; breast neoplasm; melanoma;
KW Searay syndrome; Gaucher's disease; neurological diseases;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW cardiac arrest; tachycardia; angina; infection; corneal infections;
KW wound healing; immunogen; gene therapy; antisense; food additive.
XX
XX Homo sapiens.
OS
XX
XX WO20015368-A1.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01348.
PF
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
```



CC The invention relates to isolated nucleic acid molecules encoding  
CC novel human secreted extracellular matrix proteins (SPs). The  
CC polynucleotides and proteins are used to prevent, treat a medical  
CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,  
CC chickens or sheep. For example, disorders associated with decreased  
CC expression of SPs. The SP polynucleotide or a vector expressing them may  
CC be administered to treat diseases by gene therapy. Antisense molecules  
CC may be administered to down regulate expression of SPs by binding with  
CC the cells own genes and preventing their expression. The polynucleotides  
CC may also be used as DNA probes in diagnostic assays. The SPs may also be  
CC used as antigens to produce antibodies and to identify modulators  
CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and  
CC SP and as diagnostic agents for detecting the presence of SPs in samples.  
CC The disorders include for example: immune/autoimmune diseases (e.g. HIV  
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis  
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
CC Parkinson's disease), cardio-/cerebrovascular disorders (e.g. cardiac  
CC arrest, tachycardia and angina), infections caused by bacteria, viruses  
CC and fungi and ocular disorders (e.g. corneal infections). Other uses  
CC include wound healing, maintenance of organs before transplantation.

Query Match 83.2%; Score 15.8; DB 22; Length 1183;  
Best Local Similarity 89.5%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CTGGGGCGACTTATTCGAC 19  
||||||| ||||| |||  
DB 743 CTGGGGCGCTTATTCGAC 725

RESULT 7  
AAS34773/c  
ID AAS34773 standard; CDNA; 1183 BP.

XX AAS34773;  
DT 04-DEC-2001 (first entry)

DE cDNA encoding novel human neoplastic disease associated polypeptide #7.  
XX  
XX Human; neoplastic disease associated polypeptide; cancer; gene therapy;  
KW hyperproliferative disorder; neural disorder; immune system disorder;  
KW muscular disorder; reproductive disorder; gastrointestinal disorder;  
KW pulmonary disorder; cardiovascular disorder; renal disorder;  
KW neuroprotective; cytosolic; anti inflammatory; vasotropic; ss.

OS Homo sapiens.  
XX  
XX WO200155163-A1.

PD 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01358.

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226881.  
PR 22-AUG-2000; 2000US-0226886.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.







XX 17-JAN-2001; 2001WO-US01309.  
PF  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 14-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 26-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
  
(HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
PI  
XX  
PI  
XX  
DR  
Rosen CA, Barash SC, Ruben SM;  
MPI: 2001-488781/53.  
P-PSDB; AAM43598.

XX New isolated nucleic acids and polypeptides, useful for diagnosing,  
PT treating and/or preventing human diseases and disorders -  
XX  
PS Claim 1; SEQ ID NO 112; 664bp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AA163803-AA164012) and  
CC the encoded proteins (AA1634497-AA163660) useful for preventing, treating  
CC or ameliorating medical conditions e.g. by protein or gene therapy. The  
CC genes were isolated from a range of human tissues disclosed in the  
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists  
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,  
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,  
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or  
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,  
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pft\_sequences.  
XX  
SQ Sequence 2613 BP; 558 A; 670 C; 768 G; 614 T; 3 other;  
Query Match 83.2%; Score 15.8; DB 22; Length 2613;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CTGGGGGCGACTTATTGCAC 19  
Db 2178 CTGGGGGCGCTTATTCCAC 2160  
||||||| ||||| |||  
RESULT 11  
AAC77382/c  
ID- AAC77382 standard; cDNA; 3463 BP.  
XX  
AC AAC77382;  
XX  
XX 08-FEB-2001 (first entry)  
DE Human ORFX ORF2937 polynucleotide sequence SEQ ID NO:5873.  
XX  
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antiporiatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
KW thrombosis; contraceptive; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
PN NO200058473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000; 2000WO-US08621.  
XX  
PR 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.

XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Leach M;  
XX  
DR WPI: 2000-602362/57.  
DR P-PSDB; AAB43173.  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
PS Claim 5; Page 5042-5044; 5507bp; English.  
XX  
CC AAC77446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antiporiatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressive;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy.  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 3463 BP; 721 A; 915 C; 1073 G; 753 T; 1 other;  
Query Match 83.2%; Score 15.8; DB 21; Length 3463;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CTGGGGGCGACTTATTGCAC 19  
Db 2972 CTGGGGGCGCTTATTCCAC 2954  
||||||| ||||| |||  
RESULT 12  
AAS68056  
ID AAS68056 standard; cDNA; 3531 BP.  
XX  
AC AAS68056;  
XX  
XX 13-FEB-2002 (first entry)  
DE DNA encoding novel human diagnostic protein #3860.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX



```

PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer
XX
XX Claim 1; Page 2090; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 164 BP; 35 A; 40 C; 42 G; 47 T; 0 other;
SQ
-Query Match 81.1%; Score 15.4; DB 23; Length 164;
Best Local Similarity 94.1%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 CTGGGGGCGAGTTATTGC 17
DB 141 CTGGGGGCGAGTTATTGC 157

```

```

XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer
XX
XX Claim 1; Page 7136; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 427 BP; 86 A; 109 C; 115 G; 117 T; 0 other;
SQ
-Query Match 81.1%; Score 15.4; DB 23; Length 427;
Best Local Similarity 94.1%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 CTGGGGGCGAGTTATTGC 17
DB 174 CTGGGGGCGAGTTATTGC 190

```

Search completed: July 1, 2003, 06:34:04  
Job time : 106.231 secs

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
Schlegel R, Endege WO, Monahan JE;  
PI



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 06:21:32 ; Search time 25.333 Seconds  
(Without alignments)  
230.008 Million cell updates/sec

Title: US-10-053-662A-30

Perfect score: 19

Sequence: 1 ctgggggcagttatgcac 19

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*

2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*

3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*

4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*

5: /cgn2\_6/ptodata/1/ina/PTCUTS.COMB.seq.\*

6: /cgn2\_6/ptodata/1/ina/Backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.4	81.1	4316	1	US-08-317-450B-14
2	15.4	81.1	4316	3	US-08-800-593-14
3	15.4	81.1	5156	4	US-09-643-597-130
4	15.4	81.1	5200	1	US-08-317-450B-12
5	15.4	81.1	5200	3	US-08-800-593-12
6	15.4	81.1	11443	4	US-08-961-527-49
7	14.8	77.9	2085	4	US-09-352-990-9
8	14.8	77.9	5406	4	US-08-961-527-166
9	14.4	75.8	21	3	US-09-009-913-127
10	14.4	75.8	852	4	US-09-134-001C-2815
11	14.4	75.8	1080	4	US-09-134-001C-891
12	14.4	75.8	1958	4	US-08-665-034A-3
13	14.4	75.8	2200	2	US-08-626-685A-9
14	14.4	75.8	2200	4	US-08-993-088A-1
15	14.4	75.8	2200	4	US-08-993-424B-1
16	14.4	75.8	4403765	4	US-09-103-840A-2
17	14.4	75.8	4411529	4	US-09-103-840A-1
18	14.2	74.7	39	3	US-09-289-751-2
19	14.2	74.7	40	3	US-09-289-751-1
20	14.2	74.7	1494	4	US-09-071-035-341
21	14.2	73.7	3345	1	US-07-972-791-7
22	14.2	73.7	3361	1	US-07-972-791-6
23	13.8	72.6	139	4	US-09-141-027-11
24	13.8	72.6	465	4	US-09-134-001C-1373
25	13.8	72.6	1145	4	US-09-227-794-1
26	13.8	72.6	1380	2	US-08-748-947A-1
27	13.8	72.6	2520	3	US-08-968-563-10

28	13.8	72.6	2520	3	US-08-969-683A-10	Sequence 10, Appl
29	13.8	72.6	2520	4	US-09-297-928-6	Sequence 6, Appl
30	13.8	72.6	4104	4	US-08-961-527-169	Sequence 169, App
31	13.8	72.6	6012	1	US-08-375-709-14	Sequence 14, Appl
32	13.8	72.6	6012	1	US-08-752-929-14	Sequence 23, Appl
33	13.8	72.6	7147	4	US-08-961-527-23	Sequence 69, Appl
34	13.8	72.6	8898	4	US-08-961-527-69	Sequence 1, Appl
35	13.8	72.6	37895	1	US-08-375-709-1	Sequence 1, Appl
36	13.8	72.6	37895	1	US-08-752-929-1	Sequence 1, Appl
37	13.8	72.6	37895	3	US-09-090-793-12	Sequence 12, Appl
38	13.8	72.6	40138	3	US-09-090-793-12	Sequence 163, App
39	13.4	70.5	773	4	US-08-936-165A-163	Sequence 8, Appl
40	13.4	70.5	2234	4	US-08-993-088A-8	Sequence 8, Appl
41	13.4	70.5	2234	4	US-08-993-424B-8	Sequence 1, Appl
42	13.4	70.5	5631	4	US-09-052-469-1	Sequence 1, Appl
43	13.4	70.5	12912	2	US-08-460-751-1	Sequence 5, Appl
44	13.4	70.5	13807	4	US-09-052-469-5	Sequence 4, Appl
45	13.4	70.5	14060	3	US-08-658-136-4	

## ALIGNMENTS

RESULT 1  
US-08-317-450B-14/C  
; Sequence 14, Application US/08317450B  
; Patent No. 5660982  
; GENERAL INFORMATION:  
; APPLICANT: Tryggvason, Karl  
; APPLICANT: Kallunki, Pekka  
; APPLICANT: Pyke, Charles  
; TITLE OF INVENTION: Laminin Chains: Diagnostic and  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BANNER & ALLEGRETTI, LTD.  
; STREET: Ten South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/317,450B  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Chao, Mark  
; REGISTRATION NUMBER: 37,293  
; REFERENCE/DOCKET NUMBER: 94,778  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4316 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 118..183  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 118..3453  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 4021..4316





```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,450B
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Chao, Maix
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 94,778
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 5200 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3699
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4433
FEATURE:
NAME/KEY: polyA_site
LOCATION: 5195

```

LENGTH: 11443 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-49

Query Match 81.1%; Score 15.4; DB 4; Length 11443;  
Best Local Similarity 94.1%; Pred. No. 38;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGGGGGCGAGTTATTGCA 18  
|||||  
DB 7079 TGGGGGCGAGTTATTGCA 7095

RESULT 7  
US-09-352-990-9

Sequence 9, Application US/09352990  
Patent No. 6235090  
GENERAL INFORMATION:  
APPLICANT: Famodu, Layo O.  
APPLICANT: Orozco, Buddy  
APPLICANT: Rafalski, Antoni  
TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase  
FILE REFERENCE: BB-1191  
CURRENT APPLICATION NUMBER: US/09/352,990  
CURRENT FILING DATE: 1999-07-14  
EARLIER APPLICATION NUMBER: 60/092,866  
EARLIER FILING DATE: July 15, 1998  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 9  
LENGTH: 2085  
TYPE: DNA  
ORGANISM: Zea mays  
US-09-352-990-9

Query Match 77.9%; Score 14.8; DB 4; Length 2085;  
Best Local Similarity 88.9%; Pred. No. 60;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGGCGAGTTATTGCA 18  
|||||  
DB 1926 CTGGGGGCGAGTTATTGCA 1943

RESULT 8  
US-08-961-527-166/c

Sequence 166, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527

FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 166:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5406 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-166

Query Match

Best Local Similarity 77.9%; Score 14.8; DB 4; Length 5406;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TGGGGGCGAGTTATTGCA 19  
|||||  
DB 4167 TGGGGGCGAGTTATTGCA 4150

RESULT 9  
US-09-009-913-127

Sequence 127, Application US/09009913  
Patent No. 6087485  
GENERAL INFORMATION:  
APPLICANT: Axxis Pharmaceuticals, Inc.  
TITLE OF INVENTION: Asthma Related Genes  
NUMBER OF SEQUENCES: 339  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bozicevic & Reed, LLP  
STREET: 285 Hamilton Ave, Suite 200  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94301

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/009,913  
FILING DATE: 21-JAN-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sherwood, Pamela J  
REGISTRATION NUMBER: 36,677  
REFERENCE/DOCKET NUMBER: SEO-4P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-327-3231  
TELEFAX: 650-327-3231

TELEX:  
INFORMATION FOR SEQ ID NO: 127:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Other  
US-09-009-913-127

Query Match

Best Local Similarity 75.8%; Score 14.4; DB 3; Length 21;  
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGGCGAGTTATTGCA 18

Db 2 CTGGGAGCAGTATTGCA 19

## RESULT 10

US-09-134-001C-2815  
; Sequence 2815, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2815  
; LENGTH: 852  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2815

Query Match 75.8%; Score 14.4; DB 4; Length 852;

Best Local Similarity 93.8%; Pred. No. 85;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGGGGGCGATTATTGC 17

Db 341 TGGGGGCGACTTATTGC 356

## RESULT 11

US-09-134-001C-891  
; Sequence 891, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 891  
; LENGTH: 1080  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-891

Query Match 75.8%; Score 14.4; DB 4; Length 1080;

Best Local Similarity 93.8%; Pred. No. 88;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGGGCGATTATTGCAC 19

Db 85 GGGGCGATTATTGCAC 100

## RESULT 12

US-08-665-034A-3/C  
; Sequence 3, Application US/08665034A  
; Patent No. 6410686  
; GENERAL INFORMATION:  
; APPLICANT: Bloomquist, Brian T.; McCaleb, Michael L.;  
; APPLICANT: Cornfield, Linda J.; Yoo-Warren, Heeja

TITLE OF INVENTION: Galanin Receptor and Nucleic Acid  
; TITLE OF INVENTION: Sequences  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bayer Corporation  
; STREET: 400 Morgan Lane  
; CITY: West Haven  
; STATE: CT  
; COUNTRY: US

ZIP: 06516

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS v. 6.30

SOFTWARE: Wordperfect for Windows version 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/665,034A

FILING DATE: 5-JUN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Brewer, Alice A.

REGISTRATION NUMBER: 32888

REFERENCE/DOCKET NUMBER: WH 5012-019.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (203) 812-2705

TELEFAX: (203) 812-5492

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1958

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

IMMEDIATE SOURCE:

CLONE: Clone Y107 nucleic acid sequence

US-08-665-034A-3

Query Match 75.8%; Score 14.4; DB 4; Length 1958;

Best Local Similarity 93.8%; Pred. No. 97;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTGGGGGCGATTATTG 16

Db 724 CTGGGGGCGACTGATTG 709

## RESULT 13

US-08-626-685A-9/C  
; Sequence 9, Application US/08626685A  
; Patent No. 5972624  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Kelli E.  
; APPLICANT: Gerald, Christophe P.G.  
; APPLICANT: Weishank, Richard L.  
; APPLICANT: Linemeyer, David  
; APPLICANT: Branchek, Theresa  
; APPLICANT: Fortay, Carlos  
; TITLE OF INVENTION: DNA ENCODING A MAMMALIAN  
; TITLE OF INVENTION: HYPOTHALAMIC GALANIN RECEPTOR  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/626,685A

FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 50233-B/JPW/MAT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2200 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 46..414  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1422..2171  
US-08-626-685A-9

Query Match . 75.8%; Score 14.4; DB 2; Length 2200;  
Best Local Similarity 93.8%; Pred. No. 99;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CTGGGGGCGAGTATTG 16  
|||||  
Db - 1271 CTGGGGGCGAGTATTG 1256

RESULT 14  
US-08-993-088A-1/c  
Sequence 1, Application US/08993088A  
Patent No. 6287855  
GENERAL INFORMATION:  
APPLICANT: Tan, Carina  
APPLICANT: Sullivan, Kathleen  
TITLE OF INVENTION: GALANIN RECEPTOR GALT2 AND  
TITLE OF INVENTION: NUCLEOTIDES ENCODING SAME  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merck & Co., Inc.  
STREET: P.O. Box 2000, 126 E. Lincoln Ave.  
CITY: Rahway  
STATE: NJ  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/993,088A  
FILING DATE: 18-DEC-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/033,851  
FILING DATE: 27-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Heber, Sheldon O.  
REGISTRATION NUMBER: 38,179  
REFERENCE/DOCKET NUMBER: 19846  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 732-594-1958  
TELEFAX: 732-594-4720  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2200 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-993-088A-1

Query Match . 75.8%; Score 14.4; DB 4; Length 2200;  
Best Local Similarity 93.8%; Pred. No. 99;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CTGGGGGCGAGTATTG 16  
|||||  
Db 1004 CTGGGGGCGAGTATTG 989

RESULT 15  
US-08-993-424B-1/c  
Sequence 1, Application US/08993424B  
Patent No. 6337206  
GENERAL INFORMATION:  
APPLICANT: Tan, Carina  
APPLICANT: Kolakowski, Lee F., Jr.  
TITLE OF INVENTION: MOUSE GALANIN RECEPTOR GALT2 AND  
TITLE OF INVENTION: NUCLEOTIDES ENCODING SAME  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merck & Co., Inc.  
STREET: P.O. Box 2000, 126 E. Lincoln Ave.  
CITY: Rahway  
STATE: NJ  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/993,424B  
FILING DATE: 18-DEC-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/033,851  
FILING DATE: 27-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Heber, Sheldon O.  
REGISTRATION NUMBER: 38,179  
REFERENCE/DOCKET NUMBER: 19846NP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 732-594-1958  
TELEFAX: 732-594-4720  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2200 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-993-424B-1

Query Match . 75.8%; Score 14.4; DB 4; Length 2200;  
Best Local Similarity 93.8%; Pred. No. 99;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CTGGGGGCGAGTATTG 16  
|||||  
Db 1004 CTGGGGGCGAGTATTG 989

Search completed: July 1, 2003, 07:24:00  
Job time : 29.3333 secs



```

; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1183
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (72)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-103-313-17

Query Match      83.2%; Score 15.8; DB 9; Length 1183;
Best Local Similarity 89.5%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 CTGGGGGCGAGTTATTGCAC 19
DB      743 CTGGGGGCGAGTTATTGCAC 725

RESULT 3
US-09-764-870-91/c
; Sequence 91, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 91
; LENGTH: 1183
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (72)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-870-91

Query Match      83.2%; Score 15.8; DB 10; Length 1183;
Best Local Similarity 89.5%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 CTGGGGGCGAGTTATTGCAC 19
DB      743 CTGGGGGCGAGTTATTGCAC 725

RESULT 4
US-09-764-853-50/c
; Sequence 50, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 50
; LENGTH: 1183
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
```

```

; LOCATION: (72)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-853-50

Query Match      83.2%; Score 15.8; DB 10; Length 1183;
Best Local Similarity 89.5%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 CTGGGGGCGAGTTATTGCAC 19
DB      743 CTGGGGGCGAGTTATTGCAC 725

RESULT 5
US-09-998-598-1920/c
; Sequence 1920, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madeline Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corlax Invention Disclosure Database
; SEQ ID NO 1920
; LENGTH: 129
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1920

Query Match      81.1%; Score 15.4; DB 10; Length 129;
Best Local Similarity 94.1%; Pred. No. 96;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      3 GGGGCGAGTTATTGCAC 19
DB      49 GAGGCGAGTTATTGCAC 33

RESULT 6
US-10-227-738-14/c
; Sequence 14, Application US/10227738
; Publication No. US20030100529A1
; GENERAL INFORMATION:
; APPLICANT: Trygvaason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; THERAPEUTIC USE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; APPLICATION NUMBER: US/10/227,738
; FILING DATE: 26-Aug-2002
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,593
```

```

      FILING DATE: 18-FEB-1997
      APPLICATION NUMBER: US 08/317,450
      FILING DATE: 04-OCT-1994
      ATTORNEY/AGENT INFORMATION:
        NAME: Chao, Mark
        REGISTRATION NUMBER: 37,293
        REFERENCE/DOCKET NUMBER: 94,778-B
        TELECOMMUNICATION INFORMATION:
          TELEPHONE: 312-913-0001
          TELEFAX: 312-913-0002
        INFORMATION FOR SEQ ID NO: 14:
          SEQUENCE CHARACTERISTICS:
            LENGTH: 4316 base pairs
            TYPE: nucleic acid
            STRANDEDNESS: single
            TOPOLOGY: linear
          MOLECULE TYPE: cdna
          FEATURE:
            NAME/KEY: sig_peptide
            LOCATION: 118..183
          FEATURE:
            NAME/KEY: CDS
            LOCATION: 118..3453
          FEATURE:
            NAME/KEY: repeat_unit
            LOCATION: 4021..4316
          OTHER INFORMATION: /rpt_type="other"
            /rpt_family="HUMAN ALU"
          FEATURE:
            NAME/KEY: polyA_site
            LOCATION: 4296
          SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-227-738-14

Query Match      81.1%; Score 15.4; DB 9; Length 4316;
Best Local Similarity 94.1%; Pred. NO.1.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      3 GGGGGCAGTATTGCAC 19
      1 |||||||
Db      1573 GAGGGCAGTATTGCAC 1557

      . RESULT 7
      US-09-756-071B-14/C
      ; Sequence 14, Application US/09756071B
      ; Patent No. US20020052307A1
      GENERAL INFORMATION:
        APPLICANT: tryggvason, Karl
          Pyke, Charles
          Kallunki, Pekka
        TITLE OF INVENTION: Laminin Chains: Diagnostic Uses
        NUMBER OF SEQUENCES: 20
        CORRESPONDENCE ADDRESS:
          ADDRESSEE: Fay Sharpe Fagan Minnich & McKee
          STREET: 1100 Superior Ave, Suite 700
          CITY: Cleveland
          STATE: Ohio
          COUNTRY: USA
          ZIP: 44114
        COMPUTER READABLE FORM:
          MEDIUM TYPE: floppy disk
          COMPUTER: IBM PC compatible
          OPERATING SYSTEM: PC-DOS/MS-DOS
          SOFTWARE: Patent Release #1.0, Version #1.30
        CURRENT APPLICATION DATA:
          APPLICATION NUMBER: US/09/756,071B
          FILING DATE: 08-Jan-2001
          CLASSIFICATION: <Unknown>
        PRIOR APPLICATION DATA:
          APPLICATION NUMBER: US 09/663,147
          FILING DATE: 150-September 2000
        ATTORNEY/AGENT INFORMATION:

```

```

: NAME: Minnich, Richard, J.
: REGISTRATION NUMBER: 24,175
: REFERENCE/DOCKET NUMBER: TRV 20014
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 216-861-5582
: TELEFAX: 216-241-1666
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4316 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 118..183
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 118..3453
: FEATURE:
: NAME/KEY: repeat_unit
: LOCATION: 4021..4316
: OTHER INFORMATION: /rpl_type="other"
: /rpl_family="HUMAN ALU"
: FEATURE:
: NAME/KEY: polyA_site
: LOCATION: 4296
: SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-756-071B-14

Query Match 81.1%; Score 15.4; DB 10; Length 4316;
Best Local Similarity 94.1%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0.

Oy 3 GGGGCGAGTATGCGAC 19
| | | | | | | | | | | | | | | | | | | | |
Db 1573 GAGGCGACTATGTGCAC 1557

RESULT 8
US-09-735-705-130/c
: Sequence 130. Application US/09735705
: Patent No. US20020052329A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Tonglong
: APPLICANT: Fan, Liqun
: APPLICANT: Kalos, Michael D.
: APPLICANT: Bangur, Chaitanya S.
: APPLICANT: Hosken, Nancy
: APPLICANT: Fanger, Gary R.
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aijun
: APPLICANT: Skelky, Yasir A.W.
: APPLICANT: Henderson, Robert A.
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Fanger, Neil
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
: FILE REFERENCE: 210121.455C14
: CURRENT APPLICATION NUMBER: US/09/735,705
: NUMBER OF SEQ ID NOS: 419
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 130
: LENGTH: 5156
: TYPE: DNA
: ORGANISM: Homo sapien
US-09-735-705-130

Query Match 81.1%; Score 15.4; DB 10; Length 5156;
Best Local Similarity 94.1%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0

```

```

ay      3 GGGGGCAGTTATTGCAC 19
          | |||||
Db     1521 GAGGGCAGTTATTGCAC 1505

```

```

RESULT 9
US-09-850-716A-130/c
: Sequence 130, Application US/09850716A
: Patent No. US20020115139A1
: GENERAL INFORMATION:
:
: APPLICANT: Kalos, Michael D.
: APPLICANT: McNeill, Patricia D.
:
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: OF LUNG CANCER
: FILE REFERENCE: 210121.455C15
: CURRENT APPLICATION NUMBER: US/09/850,716A
: CURRENT FILING DATE: 2001-05-07
: NUMBER OF SEQ ID NOS: 440
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 130
:
: LENGTH: 5156
:
: TYPE: DNA
: ORGANISM: Homo sapien
:
: US-09-850-716A-130

```

Query Match	81.1%;	Score 15.4;	DB 10;	Length 5156;
Best Local Similarity	94.1%;	Pred. No. 1.5e+02;		
Matches 16; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

```

OY      3 GGGGCGATTATTGCAC 19
      | | | | | | | | | |
Db      1521 GAGGCGAGTTATTGCAC 1505

```

```

RESULT 10
US-09-897-778-130/c
Sequence 130, Application US/09897778
Patent No. US20020147143n1
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Marnierakis, Margarita
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Matanabe, Yoshinhiro
APPLICANT: Henderson, Robert A.
APPLICANT: Peckham, David W.
APPLICANT: Fanger, Neil
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
OF LUNG CANCER
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.455C16
CURRENT APPLICATION NUMBER: US/09/897,778
CURRENT FILING DATE: 2001-06-28
NUMBER OF SEQ ID NOS: 467
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 130
LENGTH: 5156
TYPE: DNA
ORGANISM: Homo sapiens
US-09-897-778-130

```

Query Match	81.1%;	Score 15.4;	DB 10;	Length 5156;
Best Local Similarity	94.1%;	Pred. NO. 1.5e+02;		
Matches 16; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

```
QY      3 GGGGCAGTTATTCAC 19
          | | | | | | | | | |
Db     1521 GAGGCGAGTTATTCAC 1505
```

RESULT 11  
US-10-171-311-114/c

```

Sequence 114, Application US/10171311
Publication No. US20030087270A1
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Chen, Yan
APPLICANT: Zhao, Xumei
APPLICANT: Monahan, John
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Glatf, Karen
APPLICANT: Gannavarapu, Manjula
APPLICANT: Hoersh, Sebastian
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
FILE OF INVENTION: OF CERVICAL CANCER
FILE REFERENCE: MRI-035
CURRENT APPLICATION NUMBER: US/10/171,311
CURRENT FILING DATE: 2002-06-12
PRIORITY APPLICATION NUMBER: US 60/298,159
PRIORITY FILING DATE: 2001-06-13
PRIORITY APPLICATION NUMBER: US 60/298,155
PRIORITY FILING DATE: 2001-06-13
PRIORITY APPLICATION NUMBER: US 60/335,936
PRIORITY FILING DATE: 2001-11-14
NUMBER OF SEQ ID NOS: 238
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 114
LENGTH: 5175
TYPE: DNA
ORGANISM: Homo sapiens
US-10-171-311-114

```

Query Match	81.1%	Score 15.4;	DB 9;	Length 5175;
Best Local Similarity	94.1%;	Pred. No. 1.5e+02;		
Matches 16;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

```

QY      3 GGGGCAGTTATTGCAC 19
          | |||||
Db      1545 GAGGCAGTTATTGCAC 1529

```

RESULT 12.  
US-10-227-738-12/c  
Sequence 12, Application US/10227738  
Publication No. US20030100529A1  
GENERAL INFORMATION:  
APPLICANT: Tryggvason, Karl  
Kallunki, Pekka  
Pyke, Charles  
TITLE OF INVENTION: Laminin Chains: Diagnostic and  
Therapeutic Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
STREET: 300 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/227,738  
FILING DATE: 26-Aug-2002  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,593  
FILING DATE: 18-FEB-1997  
APPLICATION NUMBER: US 08/317,450  
FILING DATE: 04-OCT-1994  
ATTORNEY/AGENT INFORMATION:



```

; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 94,778-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 118..183
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 118..3699
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 4433
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 5195
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-227-738-12

Query Match      81.1%; Score 15.4; DB 9; Length 5200;
Best Local Similarity 94.1%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 GGGGCGAGTTATTCAC 19
Db      1573 GAGGCGAGTTATTCAC 1557

RESULT 13
US-09-756-071B-12/c
; Sequence 12, Application US/09756071B
; Patent No. US20020052307A1
; GENERAL INFORMATION:
; APPLICANT: Trygsvason, Karl
;              Kallunki, Pekka
;              Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay Sharpe Fagan Minnich & McKee
; STREET: 1100 Superior Ave, Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: USA
; ZIP: 44114
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756,071B
; FILING DATE: 08-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/663,147
; FILING DATE: 150-September 2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard, J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: TRV 20014
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 216-861-5582
; TELEFAX: 216-241-1666

```

```

; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 118..183
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 118..3699
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 4433
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 5195
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-756-071B-12

Query Match      81.1%; Score 15.4; DB 10; Length 5200;
Best Local Similarity 94.1%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 GGGGCGAGTTATTCAC 19
Db      1573 GAGGCGAGTTATTCAC 1557

RESULT 14
US-10-044-090-558/c
; Sequence 558, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO: 558
; LENGTH: 5460
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 149791.5
US-10-044-090-558

Query Match      81.1%; Score 15.4; DB 12; Length 5460;
Best Local Similarity 94.1%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 GGGGCGAGTTATTCAC 19
Db      1575 GAGGCGAGTTATTCAC 1559

RESULT 15
US-10-092-154-1809
; Sequence 1809, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See file wrapper or Palm

```

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1809  
; LENGTH: 9542  
; TYPE: DNA  
; ORGANISM: HOMO sapiens  
US-10-092-154-1809

Query Match 81.1%; Score 15.4; DB 9; Length 9542;  
Best Local Similarity 94.1%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGGGGGCGACTTATTGCA 18  
|||||  
DB 1525 TGGGGGCGACTTATTGCA 1541

Search completed: July 1, 2003, 07:26:19  
Job time : 65.8205 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 03:57:39 ; Search time 859.385 Seconds  
(without alignments)  
358.063 Million cell updates/sec

Title: US-10-053-662a-30

Perfect score: 19

Sequence: 1 ctgggggcagctatgcac 19

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_estbta:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estm:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.4	91.6	1201	13	BM046418 603626447
2	16.4	86.3	144	9	AL842103
3	16.4	86.3	183	9	AL194150
4	16.4	86.3	271	14	M78847
5	16.4	86.3	396	12	BF850108
6	16.4	86.3	457	10	AM134919

Result No.	Score	Query Match	Length	DB ID	Description
7	16.4	86.3	458	13	BC999180 PM0-RT131
8	16.4	86.3	489	9	AL048818 DKF2434F
9	16.4	86.3	524	13	BI076955
10	16.4	86.3	544	10	AM655709
11	16.4	86.3	701	17	BH428192
12	16.4	86.3	701	17	AG132023
13	16.4	86.3	716	13	BU153383
14	16.4	86.3	752	12	BC282508
15	16.4	86.3	866	17	BH442712
16	16.4	86.3	984	17	BC282706
17	16.4	86.3	1101	17	CNS05M09
18	16.4	86.3	1298	14	BQ049927
19	16.4	86.3	1469	10	BE130964
20	16.4	86.3	1562	17	AC065551
21	16.4	86.3	1588	12	BE857007
22	16.4	86.3	1588	9	AI858860
23	16.4	86.3	1588	14	T36249
24	16.4	86.3	1588	12	BF916711
25	16.4	86.3	1588	12	BF916711
26	16.4	86.3	1588	13	BI033206
27	16.4	86.3	1588	13	AM801445
28	16.4	86.3	1588	10	AM801444
29	16.4	86.3	1588	13	BI16726
30	16.4	86.3	1588	12	BF482335
31	16.4	86.3	1588	10	AM778993
32	16.4	86.3	1588	10	AM514348
33	16.4	86.3	1588	9	AA742413
34	16.4	86.3	1588	10	AM801448
35	16.4	86.3	1588	9	AA398958
36	16.4	86.3	1588	9	AI924185
37	16.4	86.3	1588	9	AA939340
38	16.4	86.3	1588	17	AQ278480
39	16.4	86.3	1588	9	AI337830
40	16.4	86.3	1588	9	AA703508
41	16.4	86.3	1588	12	BE749722
42	16.4	86.3	1588	14	BQ319898
43	16.4	86.3	1588	12	BF889999
44	16.4	86.3	1588	9	AA459604
45	16.4	86.3	1588	12	BF890005

## ALIGNMENTS

RESULT 1  
LOCUS BM046418 1201 bp mRNA linear EST 07-NOV-2001  
DEFINITION 603626447F1 NIH\_MGC\_40 Homo sapiens CDNA clone IMAGE:5453002 5',  
ACCESSION BM046418  
VERSION BM046418.1 GI:16775685  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1201)  
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCTD/DDP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1CM1945 row: n column: 11  
High quality sequence start: 3  
High quality sequence stop: 117.  
Location/Qualifiers

source  
1. .1201  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5453002"  
/clone\_1ib="NIH\_MGC\_40"  
/tissue\_type="carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: prostate; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH-MGC library."

BASE COUNT 358 a 416 c 239 g 187 t 1 others

Query Match 91.6%; Score 17.4; DB 13; Length 1201;  
Best Local Similarity 94.7%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTGGGGCAGTTATTCAC 19  
|||||

DB 700 CTGGGGCAGTTATTCAC 718

RESULT 2  
AL842103 144 bp mRNA linear EST 30-JUL-2002  
LOCUS AL842103 F000F Takifugu rubripes cDNA clone F000F15a1, mRNA  
DEFINITION sequence.  
ACCESSION AL842103  
VERSION AL842103.1 GI:22019931  
KEYWORDS EST.  
SOURCE Takifugu rubripes.  
ORGANISM Takifugu rubripes.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Takifugu.  
1 (bases 1 to 144)  
Clark, M.S.  
Takifugu rubripes ESTs  
Unpublished (2002)  
Contact: Clark MS  
MRC Human Genome Mapping Project Resource Centre  
Hinxton, Cambridge, CB10 1SB, UK  
Email: biohelp@hmp.mrc.ac.uk  
Vector: pME185-FL3  
V\_Type: phagemid  
PRIMER: ME-735FW  
Library created by Koichi Kawakami, Masahide Sasaki, Yutaka Suzuki,  
Sumio Sugano  
The Institute of Medical Science, The University of Tokyo,  
Shirokanedai, Minato-ku, Tokyo 108-8639, Japan  
and  
Kiyoshi Kikuchi, Shugo Watabe  
Laboratory of Aquatic Molecular Biology and Biotechnology, Graduate  
School of Agricultural and Life Sciences, The University of Tokyo,  
Bunkyo-ku, Tokyo 108-8639, Japan  
Library sequenced by Melody S. Clark and Amanda Thompson MRC Human  
Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10  
1SB, UK.

FEATURES  
source  
1. .144  
/organism="Takifugu rubripes"  
/db\_xref="taxon:31033"  
/clone="F000F15a1"  
/clone\_1ib="F000F"  
/sex="female"  
/tissue\_type="fin"  
/dev\_stage="adult"

BASE COUNT 37 a 40 c 38 g 29 t  
ORIGIN  
Query Match 86.3%; Score 16.4; DB 9; Length 144;  
Best Local Similarity 94.4%; Pred. No. 4.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTGGGGCAGTTATTCAC 18  
|||||

DB 116 CTGGGGCAGTTATTCAC 133

RESULT 3  
A1194150 183 bp mRNA linear EST 13-OCT-1998  
LOCUS ue75H02.r1 Soares\_NMPu Mus musculus cDNA clone IMAGE:1496979 5'  
DEFINITION similar to gb:X71642 M.musculus G8C-154 mRNA (MOUSE);, mRNA  
sequence.  
ACCESSION A1194150  
VERSION A1194150.1 GI:3745357  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 183)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:934583  
Trace considered overall poor quality  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 1.

FEATURES  
source  
1. .183  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1496979"  
/clone\_1ib="Soares\_NMPu"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; 1st strand cDNA was prepared from  
pregnant mouse uterus, and was then primed with a Not I -  
oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pT73  
vector. Library is normalized. Library was constructed by  
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 74 a 26 c 48 g 35 t  
ORIGIN  
Query Match 86.3%; Score 16.4; DB 9; Length 183;  
Best Local Similarity 94.4%; Pred. No. 4.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGGGGCAGTTATTCAC 19  
|||||

Db		58	TGGGCGAGTTATTGCAC	75	
RESULT 4					
LOCUS	M78847				
DEFINITION	M78847	271 bp	mRNA	linear	EST 26-MAY-1992
ACCESSION	EF000995	Hippocampus, Striatum gene (cat. #936205)	Homo sapiens cDNA clone HHCNH22, mRNA sequence.		
VERSION	M78847.1	GI:273162			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	(bases 1 to 271) Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M., Uterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.				
TITLE	Sequence identification of 2,375 human brain genes				
JOURNAL	Nature 355, 632-634 (1992)				
MEDLINE	92168112				
COMMENT	Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel.: 3018699056 Fax: 3018699423 Email: arkerlav@tigr.org Seq primer: M13 Forward. Location/Qualifiers 1..271				
FEATURES	/organism="Homo sapiens" /db_xref="ATCC (lnhost):78301" /db_xref="GDB:D051828E" /db_xref="taxon:9606" /clone="HHCNH22" /clone_id="Hippocampus, Striatum gene (cat. #936205)" /note=Vector: lambdaZAP-II; Female, 2 years; oligo-dT + random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb average insert size."				
BASE COUNT	54 a	98 c	57 g	60 t	2 others
ORIGIN					
Query Match	86.3%	Score 16.4;	DB 14;	Length 271;	
Best Local Similarity	89.5%	Pred. No. 4.9e+02;			
Matches 17; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;	
Oy	1 CTGGGGCAGTTATTGCAC	19			
Db					
	218 CTGGGGCTGTATTNCAC	236			
RESULT 5					
LOCUS	BF850108	396 bp	mRNA	linear	EST 16-JAN-2001
DEFINITION	KO-EN0080-161100-021-a05 EN0080	Homo sapiens cDNA, mRNA sequence.			
ACCESSION	BF850108				
VERSION	BF850108.1	GI:12237258			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	(bases 1 to 396) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)				

MEDLINE	20202663	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC0&t=RC0-EN0080- 16100-021-805&t3=2000-11-16&t4=1) Seq primer: puc 18 forward High quality sequence start: 63 High quality sequence stop: 396. location/Qualifiers 1. 396 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_id="EN0080" /dev_stage="Adult" site="Organ: lung,normal; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTS PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT	95 a 90 c 110 g 101 t	
ORIGIN		
Query Match	86.3%; Score 16.4; DB 12; Length 396;	
Best Local Similarity	94.4%; Pred. No. 5.2e+02;	
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
OY	1 CTGGGGCAGCTATTGCA 18 	
Db	247 CTGGGGCAGCTATTGCA 264	
RESULT 6		
AM134919/c		
LOCUS	AM134919 457 bp mRNA linear EST 28-OCT-1999	
DEFINITION	UI-H-B11-abr-d-05-0-UI.s1 NCI-CCAP-Sub3 Homo sapiens cDNA clone	
IMAGE:2712633 3', mRNA sequence.		
AM134919		
ACCESSION	AM134919.1 GI:6138465	
VERSION	EST.	
KEYWORDS	human.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 457) NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cga@bbs-remail.nih.gov The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA library Preparation: M.B. Soares lab clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: www-bio.llnl.gov/dbp/image/image.html Seq primer: M13 Forward POLY-A=yes.	
JOURNAL		
COMMENT		
FEATURES	location/Qualifiers 1. 457 /organism="Homo sapiens" /db_xref="taxon:9606"	
SOURCE		

/clone="IMAGE:2712633"  
 /clone.lib="NCI CGAP Sub3"  
 /lab\_host="DB10B (Life Technologies)"  
 /note/Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
 NCI CGAP Sub3 library is a substracted library derived from  
 the NCI CGAP Sub1 library, which is a substracted library  
 derived from B1. B1 constitutes a mixture of 21  
 normalized or substracted NCI CGAP libraries: NCI CGAP Co4  
 , NCI CGAP Pr22, NCI CGAP Pr28, NCI CGAP Co10,  
 NCI CGAP Co16, NCI CGAP Kid5, NCI CGAP Kid12,  
 NCI CGAP Kid3, NCI CGAP Kid11, NCI CGAP Lym2,  
 NCI CGAP Br2, NCI CGAP Co8, NCI CGAP CLU1, NCI CGAP Le12,  
 NCI CGAP Brn23, NCI CGAP Lu5, NCI CGAP Lu24,  
 NCI CGAP Lu19, NCI CGAP GC4, NCI CGAP GC6,  
 NCI CGAP Brn25. These 21 libraries were pooled and a  
 single-stranded DNA preparation of the resulting mixture  
 was used as a tracer in a subtractive hybridization with  
 a driver whose composition is detailed below:  
 NCI CGAP Kid3 pool 1 LLM 3334-3337, 3682-3683,  
 3798-3803 (IMAGE Clonoids 1322376-1323911, 1456008-1456757)  
 , 1500552-1502855); NCI CGAP Kid5 pool 1 LLM 3338-3342  
 , 3722-3725, 3776-3778 (IMAGE Clonoids 1323912-1325831,  
 1471368-1472903, 1492104-1493255); NCI CGAP Lu5 pool 1  
 LLM 3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991,  
 1520904-1522439); NCI CGAP GC4 pool 1 LLM 3164-3167,  
 3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631,  
 1469064-1470983, 1475592-1476743); NCI CGAP Pr22 pool 1  
 LLM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids  
 985608-986759, 1101192-1101959, 1217928-1220615);  
 NCI CGAP Co10 pool 1 LLM 2644-2653, 2871-2872 (IMAGE  
 Clonoids 1057416-1061255, 1144584-1145351). Subtraction  
 was performed as previously described [Bonaldi, Lennon &  
 Soares (1996)]: Normalization and Subtraction: Two  
 Approaches To Facilitate Gene Discovery. Genome Research  
 6, 791-806.  
 TAG\_LIB=NCI CGAP Co10  
 TAG\_TISSUE=colon  
 TAG\_SEQ=AAACG

BASE COUNT 130 a 88 g 122 t  
 ORIGIN

Query Match 86.3%; Score 16.4; DB 10; Length 457;  
 Best Local Similarity 94.4%; Pred. No. 5.4e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTGGGGCGACTTATGCA 18  
 ||||| ||||| ||||| |||||  
 Db 343 CTGGGGCGACTTATGCA 326

RESULT 7  
 LOCUS BG999180 458 bp mRNA linear EST 13-JUN-2001  
 DEFINITION PM0-HT1311-270401-002-B09 HT1311 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BG999180  
 VERSION BG999180.1 GI:14403252  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 458)  
 Dias Neto,E., Garcia Correa,R., Vejlovski-Almeida,S., Brites,M.R.,  
 Nagai,M.A., da Silva,W.F., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.J., Matsukuma,A., Bata,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

TITLE  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663

COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM0&ct2=PM0-ht1311-270401-002-B09&f3=2001-04-27&f4=1) Seq primer: puc 18 forward High quality sequence start: 14 High quality sequence stop: 457. Location/Qualifiers			
FEATURES	1..458			
SOURCE	/organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="HT1311" /dev_stage="Adult" /note="Organ: head,neck; Vector: puc18; Site:1; Sma1; Site:2; Sma1; A mini-library was made by cloning products derived from ORSTS PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."			
BASE COUNT	122 a	123 c	100 g	113 t
ORIGIN				
Query Match	86.3%	Score 16.4;	DB 13;	Length 458;
Best Local Similarity	94.4%;	Pred. No. 5.4e+02;		
Matches 17; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
Py	1	CTGGGGGACGATTATGCA	18	
Db	254	CTGGGGGACATTATGCCA	237	
RESULT 8	AL048818 489 bp mRNA linear EST 29-FEB-2000			
LOCUS	AL048818/c			
DEFINITION	DKFZp434f2018_r1 434 (synonym: htes3) Homo sapiens cDNA clone			
ACCESSION	AL048818			
VERSION	AL048818.1 GI:4728127			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 489)			
AUTHORS	Ottensaeider,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and Wiemann,S.			
TITLE	EST (Ottensaeider, et al.)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Ottensaeider B MPS Am Klopfersplitz 18a D-82152 Martinsried, Germany This is the 5' sequence of the clone insert Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Medigenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project..sl sequence also available. This clone (DKFZp434f2018) is available at the RZPD in Berlin. Please Contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de. Location/Qualifiers 1..489 /organism="Homo sapiens" /db_xref="taxon:9606"			
FEATURES				
SOURCE				

```

/clone="DKFZp434F2018"
/clone_lib="434 (synonym: htes3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/Note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT      139 a      114 c      109 g      127 t
ORIGIN

Query Match      86.3%; Score 16.4; DB 9; Length 489;
Best Local Similarity 94.4%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTGGGGGCGAGTTATTGCA 18
        |||||||
DB      56 CTGGGGGCGAGTTATTGCA 39

RESULT 9
BI076955      524 bp      mRNA      linear      EST 20-JUN-2001
LOCUS      NXPV_085_C02_F NXPV (Nsf xylem planings wood vertical) Pinus taeda
ACCESSION      CDNA clone NXPV_085_C02_5, mRNA sequence.
VERSION      BI076955
KEYWORDS      BI076955.1 GI:14515612
SOURCE      EST.
ORGANISM      loblolly pine.
              Pinus taeda
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
              Sederoff, R.
              Molecular Basis of Wood Formation in the Pine Megagenome
              Unpublished (2000)
              Contact: Johnson, Arthur
              North Carolina State University
              Tel: 919 515 7800
              Fax: 919 515 7801
              Email: ajohnson@unity.ncsu.edu
              Seq primer: T3.

FEATURES
    source
        1..524
        /organism="Pinus taeda"
        /strain="Coastal plain loblolly pine from North Carolina"
        /db_xref="taxon:3352"
        /clone="NXPV_085_C02"
        /clone_lib="NXPV (Nsf xylem planings wood vertical)"
        /tissue_type="xylem"
        /cell_type="planings (secondary)"
        /dev_stage="transitional"
        /lab_host="X11-Blue"
        /note="Vector: Bluescript SK; Site_1: Eco RI; Site_2: XhoI
        ; The library is from early (spring) secondary wood, taken
        from a ten year old tree in the transitional phase. The
        tree is a kind gift of the Westvaco Corporation. Secondary
        xylem was harvested from the tree by peeling back the bark
        and primary xylem and then removing the underlying tissue
        with a block plane. NOTE: The sequences contain a 'cdna
        adapter' between the EcoRI site and the start of the EST.
        The adapter sequence is 'AATTGGCAGAG'."

BASE COUNT      144 a      92 c      110 g      162 t
ORIGIN

Query Match      86.3%; Score 16.4; DB 13; Length 524;
Best Local Similarity 94.4%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTGGGGGCGAGTTATTGCA 18
        |||||||
DB      152 CTAGGGGCGAGTTATTGCA 169

RESULT 10

```

```

AM655709
LOCUS      AM655709      544 bp      mRNA      linear      EST 25-APR-2001
DEFINITION      106914 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION      AM655709
VERSION      AM655709.1 GI:7421535
KEYWORDS      EST.
SOURCE      cow.
ORGANISM      Bos taurus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
              Bovidae; Bovinae; Bos.
              1 (bases 1 to 544)
              Smith, R.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
              Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett
              G.L., Heaton, M.P., Laegreid, W.M., Rohrer, G.A., Chitko-McKown, C.G.,
              Perlea, G., Holt, T., Karameycheva, S., Liang, F., Quackenbush, J. and
              Keele, J.W.
              Sequence evaluation of four pooled-tissue normalized bovine cDNA
              libraries and construction of a gene index for cattle
              Genome Res. 11 (4), 626-630 (2001)
              21180013
              Contact: Smith RPL
              USDA, ARS, US Meat Animal Research Center
              PO Box 166, Clay Center, NE 68933-0166, USA
              Tel: 402 762 4366
              Fax: 402 762 4390
              Email: smith@email.marc.usda.gov
              Single pass sequencing. Bases called and alt-trimmed with phred
              v0.980904.e. Vector identified by cross-match with the -mnscore 18
              and -minmatch 12 options.
              PCR Primers
              FORWARD: AGGAACAGCTATGACCAT
              BACKWARD: GTTTCACAGCAGCAGC
              Plate: 80 row: 1 column: 21
              Seq primer: ATTGAGTGACACTATAG.

FEATURES
    source
        1..544
        /organism="Bos taurus"
        /db_xref="taxon:9913"
        /clone_lib="MARC 1BOV"
        /tissue_type="pooled"
        /lab_host="DH10B"
        /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
        library made from pooled tissue from lymph node, ovary,
        fat, hypothalamus, and pituitary."

BASE COUNT      136 a      142 c      142 g      124 t
ORIGIN

Query Match      86.3%; Score 16.4; DB 10; Length 544;
Best Local Similarity 94.4%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTGGGGGCGAGTTATTGCA 18
        |||||||
DB      507 CTGGGGGCGAGTTATTGCA 524

RESULT 11
BH428192      701 bp      DNA      linear      GSS 12-DEC-2001
LOCUS      BOHPV85TF BOHP Brassica oleracea genomic clone BOHPV85, DNA
DEFINITION      sequence.
ACCESSION      BH428192
VERSION      BH428192.1 GI:17613920
KEYWORDS      GSS.
SOURCE      Brassica oleracea
              Brassica oleracea
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
              1 (bases 1 to 701)
              Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.
              Whole genome shotgun sequencing of Brassica oleracea

REFERENCE
    AUTHORS      Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.
    TITLE      Whole genome shotgun sequencing of Brassica oleracea

```

JOURNAL Unpublished (2001)  
COMMENT Contact: Chris Town

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

Email: cdtown@elgr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF

Class: sheared ends.

FEATURES  
source Location/Qualifiers

1..701  
/organism="Brassica oleracea"  
/strain="TO100D03"  
/db\_xref="taxon:3712"  
/clone="BOHPV85"  
/clone.lib="BOHP"  
/note="Vector: pROSL, Site\_1: BstXI, 2-3 kb sheared  
genomic DNA inserted into pROSL using BstXI linkers"

BASE COUNT 211 a 136 c 104 g 250 t

ORIGIN

Query Match 86.3%; Score 16.4; DB 17; Length 701;  
Best Local Similarity 94.4%; Pred. No. 5.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 TGGGGGCGAGTTATTCAC 19  
|||||  
Db 301 TGGAGCGAGTTATTCAC 284

RESULT 12 AG132023 701 bp DNA linear GSS 04-NOV-2001  
LOCUS Pan troglodytes DNA, clone: PTB-144E08.F, genomic survey sequence.  
DEFINITION AG132023  
ACCESSION AG132023  
VERSION AG132023.1 GI:16661701

KEYWORDS GSS.  
SOURCE Pan troglodytes male lymphoblast DNA, clone.lib:PTB chimpanzee male

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (02-NOV-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: chimpesgsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.

PRIMERS  
Sequencing: -21M13

LIBRARY  
Vector : PKS145  
R.Site 1 : SacI  
R.Site 2 : SacI

FEATURES  
source Location/Qualifiers

1..701  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-144E08.F"  
/sex="male"  
/cell\_type="lymphoblast"

BASE COUNT 151 a 285 c 148 g 117 t  
ORIGIN

Query Match 86.3%; Score 16.4; DB 17; Length 701;  
Best Local Similarity 94.4%; Pred. No. 5.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTGGGGCGAGTTATTCGA 18  
|||||  
Db 517 CTGGGGCGAGTTATTCGA 500

RESULT 13 BJ153383 716 bp mRNA linear EST 24-JAN-2002  
LOCUS BJ153383 Unpublished oligo-capped cDNA library, C. elegans L1 stage  
DEFINITION Caenorhabditis elegans cDNA clone yk1318906 3', mRNA sequence.  
ACCESSION BJ153383  
VERSION BJ153383.1 GI:18321368

KEYWORDS

SOURCE

Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae  
; Rhabditidae; Pelodermidae; Caenorhabditis.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..716

/organism="Caenorhabditis elegans"

/strain="N2"

/db\_xref="taxon:6239"

/clone="yk1318906"

/clone.lib="unpublished oligo-capped cDNA library, C.

elegans L1 stage"

/sex="hermaphrodite"

/tissue.type="whole animal"

/dev.stage="L1"

/note="The AD-wrncDNA library was generated with poly(A)+

RNA isolated from both hermaphrodite and male N2 worms of

all larval stages, embryos, adults and dauers and the

subsequent generation of cDNAs by poly(A) priming. The

cDNAs were cloned into pPC86"

BASE COUNT 202 a 131 c 163 g 220 t

ORIGIN

Query Match 86.3%; Score 16.4; DB 13; Length 716;  
Best Local Similarity 94.4%; Pred. No. 5.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTGGGGCGAGTTATTCGA 18  
|||||  
Db 636 CTGGGGCGAGTTATTCGA 619

RESULT 14 BG282508 752 bp mRNA linear EST 21-FEB-2001

LOCUS BG282508 602406354F1 NIH\_MGC\_91 Homo sapiens cDNA clone IMAGE:4518661 5',  
DEFINITION mRNA sequence.  
ACCESSION BG282508  
VERSION BG282508.1 GI:13031617  
KEYWORDS EST.



SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 752)  
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs.rem@nih.gov  
Tissue Procurement: DCTD/DRP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10413 row: k column: 14  
High quality sequence stop: 486.  
Location/Qualifiers  
1. 752  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4518661"  
/clone\_lib="NIH-MGC\_91"  
/tissue\_type="adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: prostate; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.4 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH-MGC Library."

BASE COUNT 232 a 225 c 153 g 142 t  
ORIGIN

Query Match 86.3%; Score 16.4; DB 12; Length 752;  
Best Local Similarity 94.4%; Pred. No. 5.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTGGGGGCGACTTATTGCA 18  
|||||  
Db 476 CTGGGGGCGACTTATTGCA 459

RESULT 15  
BH442712 866 bp DNA linear GSS 12-DEC-2001  
LOCUS BOGGG28TR BOGG Brassica oleracea genomic clone BOGGG28, DNA  
DEFINITION  
sequence.  
ACCESSION BH442712  
VERSION BH442712.1 GI:17628426  
KEYWORDS GSS.  
SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 866)  
AUTHORS Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.  
TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)  
COMMENT Other GSSs: BOGGG28TR  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: 7R  
Class: sheared ends.  
FEATURES  
source 1. 866  
Location/Qualifiers

/organism="Brassica oleracea"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone="BOGGG28"  
/clone\_lib="BOGG"  
/note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared  
genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 295 a 144 c 182 g 245 t  
ORIGIN

Query Match 86.3%; Score 16.4; DB 17; Length 866;  
Best Local Similarity 94.4%; Pred. No. 6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGGGGGCGACTTATTGCA 19  
|||||  
Db 525 TGGAGGCGACTTATTGCA 542

Search completed: July 1, 2003, 07:03:46  
Job time : 864.385 secs

